siris

10 TALES OF REPRODUCIBILITY

Funded by
the European Union

Contributors
Paula Muñoz, KU Leuven
Magdalena Kozula, KU Leuven
Stefan Wouters, University Medical Center Utrecht
Ayu Dewi, Amsterdam University Medical Centers
Constant Vinatier, Université Rennes
Nicholas DeVito, University of Oxford
Veerle Van den Eynden, KU Leuven

Reviewers
Florian Naudet, Université Rennes
Mariska Leeflang, Amsterdam University Medical Centers
Inge Stegeman, University Medical Center Utrecht
Veerle Van den Eynden, KU Leuven
Magdalena Kozula, KU Leuven

Editing
Veerle Van den Eynden, KU Leuven
Hynek Roubik, Czech University of Life Sciences Prague
Stacy Hammond, Czech University of Life Sciences Prague

Design
Stacy Hammond, Czech University of Life Sciences Prague
Paula Castro, Czech University of Life Sciences Prague
Masha Gulak, Czech University of Life Sciences Prague

Generative AI tools have been used to generate images and to enhance English writing for some tales.

10 TALES OF REPRODUCIBILITY

Reproducibility is crucial to the progress and impact of research and innovation, as it confirms or corrects the outcomes of single studies. Reproducibility is understood as a continuum from the reproduction of results, based on the same data, code and methodology, to replication, meaning the repeating of results using the same analytical method, but with different datasets[1]. The ability to reproduce studies is thus key to the credibility, integrity and trustworthiness of science. When research findings can be independently verified, science moves forward with confidence. However, reproducibility remains a major challenge in many research disciplines. When studies cannot be reproduced, findings risk being misinterpreted or misapplied.

Multiple stakeholders have a vested interest in the research landscape: research funders, publishers, journals and research performing organisations. Through their policies, guidelines, initiatives, support and interventions, they can influence the quality and trustworthiness of research by removing barriers that hinder reproducibility and promoting facilitators that enhance it.

Ten Tales of Reproducibility is a collection of inspiring examples of reproducibility practices, identified through two years of research in the Open Science to Increase Reproducibility in Science (OSIRIS) project. We investigate facilitators, barriers and effective practices and interventions that improve reproducibility at the funding, publishing, university and researcher levels.

With a systematic scoping review, we assessed the extent to which Open Science interventions have been tested with rigour for their effectiveness in improving reproducibility. We interviewed 60 European researchers to examine their views, practices and motivations for reproducibility across disciplines, with factors that facilitate or hinder. During focus group discussions with research funders, journals, research institutions, reproducibility networks, reproducibiliTea journal clubs and civil society organisations, we explored roles and responsibilities of these stakeholders to assure the quality and trustworthiness of research, discussing initiatives to facilitate reproducibility or remove barriers. Using TOP 2025 criteria, we assessed how research institutions, journals, and funders address and promote reproducibility in their policies, procedures and guidelines.

From this research, ten remarkable cases of reproducibility practice by different stakeholders were selected to inspire how reproducibility in science can be supported and improved. Cases include bottom-up and top-down initiatives by research institutions, journals, funders and the research community. They convey successes, challenges, lessons learned and wider applicability.

By sharing these ten tales, we want to highlight the positive aspects of the journey, the lessons learned, and the values to uphold. These tales can contribute to promoting best practices around reproducibility, transparency and integrity of research.

[1] *European Commission: Directorate-General for Research and Innovation, Athena RC, Know-Center and PPMI (2022). Assessing the reproducibility of research results in EU Framework Programmes for Research. https://data.europa.eu/doi/10.2777/18678*

OPEN SCIENCE TO INCREASE REPRODUCIBILITY OF SCIENCE

Embedding reproducibility in the strategy and design of research should be regarded as a key precondition to research quality. While over the past decade, many interventions to improve reproducibility have been introduced, targeted at funders, publishers or individual researchers, only a few of them have been empirically tested. We need a paradigm and culture shift to reform the research and innovation system, bottom up and top down, to regain overall trust in science.

OSIRIS facilitates this shift by:

- Investigating the underlying drivers and effective interventions that increase reproducibility at the level of research funders, publishers and journals, research institutes, researchers and the research community at large
- Co-creating evidence-based solutions for researchers and institutions to increase reproducibility and testing them through randomised controlled trials with researcher networks
- Assessing whether grant proposal referees are more capable of predicting the reproducibility of research projects when using an open science checklist than without
- Assessing whether using an open science checklist during peer review of manuscripts improves the reproducibility of scientific findings in reviewed manuscripts
- Developing dashboards of indicators of reproducible research practices
- Developing guidance for funders, publishers, researchers and peer reviewers for judging reproducibility
- Based on all OSRIS findings, co-creating, testing and disseminating training resources for researchers, funders and journals to embed reproducibility in research
- Performing quality audits at the project and output levels to test these novel practices.

This is realised by an interdisciplinary team of scientists with hands-on expertise in open science, reproducibility, implementation, education and data sharing, together with committed early career researchers, publishers and funders, who are closely involved in the project through dedicated co-creation activities and as part of our advisory board.

OSIRIS is coordinated by the University Medical Center Utrecht. Partners are Amsterdam University Medical Centres, the Czech University of Life Sciences, the Hungarian University of Agriculture and Life Science, KU Leuven, the Mario Negri Institute for Pharmacological Research, the National Institute of Health and Medical Research, the University of Oxford and Sense About Science.

Follow OSIRIS on https://osiris4r.eu and https://osf.io/8nyke

siris

LIST OF TALES

Writer: *Paula Muñoz*

A RESEARCH INSTITUTION TALE

01 RESEARCH INTEGRITY SUPPORT AT THE MANCHESTER CANCER RESEARCH UK INSTITUTE

Interviewee: *Andrew Porter, Cancer Research UK Manchester Institute*

While many research institutions have research integrity officers to support their researchers, the Cancer Research UK Manchester Institute takes this one step further. Since 2021, they have had a dedicated research integrity and training advisor, Andrew Porter, whose role is to support overall good research practice across the institute. A big part of the role is the pre-submission review of research papers that are authored by researchers from the institute, to advise them on how to ensure that the research described in the paper is reproducible and trustworthy.

Andrew supports over 200 researchers at the Institute. The research ranges from basic research, like *in vitro* cell-based assays, to studies that use animal models and translational studies involving samples from clinical trials, covering a spectrum of cancer research topics.

What sets this initiative apart is the specificity and personalisation of the advice provided and the complementary training researchers receive.

> *"I review in detail each manuscript before submission to a journal, using a checklist, and make comments or suggestions for improvement. Then the authors can come see me in person for guidance on aspects they may not have the time or capacity to address."*

Manuscript reviews

For reviews, Andrew uses a checklist that can be applied to any kind of research. He looks at authorship, whether credit taxonomy is used, whether the acknowledgements are comprehensive, and that conflict of interest statements are included.

Checking data reporting is important, as almost all manuscripts have Cancer Research UK as a funder, which has clear data sharing and openness guidelines. Plus, in cancer research, much of the data may have been contributed by cancer patients. There's then a responsibility to make sure to get the best out of that data, and it doesn't just form a single data point in one study, but becomes FAIR data in a repository. So he supports researchers to share data via suitable repositories, and checks that external sources are correctly cited.

He looks specifically at ethics around animal research, making sure that the paper meets the ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) guidelines. And if not, then bespoke suggestions are given on how to improve that. For clinical trials, he checks that it has been registered and that the right ethical approvals for human studies are in place.

He checks the resolution of images, the clarity and accessibility of graphs to avoid them being confusing or biased, and checks that detailed statistical information is provided for each figure.

In the methods section, he makes sure it reflects the breadth of what was actually done, that no outdated methods are cited, and whether protocols.io could be used to deposit a protocol with a persistent identifier. When it's a very technical paper, he might flag that the writing may be quite dense for non-specialist readers. He provides advice for checking references to avoid citing retracted studies.

He uses two specific tools during checks: iThenticate to detect whether there are issues with plagiarism – that will also pick up if a methods section is repeated from a previous paper and needs citing – and Imagetwin, because a common error in papers is using the same image or graph in two different places without realising it.

Hands-on for reproducibility

One-to-one meetings give an opportunity to discuss specific questions related to research projects. His approach is to be accessible and helpful, facilitating individuals in finding the answers or resources they need.

If a new policy or guidelines emerge that may impact ongoing research projects, Andrew proactively informs researchers. If faced with more complex problems, he can seek help from external experts or direct researchers to appropriate sources of advice and support. Researchers thus get specific, continuous and personalised guidance to improve the reproducibility of their research.

He delivers training sessions to improve general competencies related to reproducibility, on topics such as data sharing, fidelity and representation. This helps researchers become more autonomous in their research practices.

Highlights and challenges

One of the highlights is being immersed in emerging and novel science, as he gets to read papers early and works with the communication team to prepare stories for when a paper comes out.

The role aims to improve the reproducibility of research at the institute through continuous and personalised support, supporting an Open Science culture that promotes collaboration, transparency and critical thinking.

A challenge can be the volume of work that could be overwhelming for a single person. Some barriers extend beyond the immediate research environment, such as the time and resources required to comply with a wide range of institutional, funder and government policies or inadequate infrastructure. Addressing these challenges effectively requires system-wide change, which exceeds what individual initiatives can achieve.

Looking ahead

One goal is to improve personalisation by creating a visual training pathway for new PhDs, postdocs and junior group leaders. He plans to develop training modules based on common issues he observes in research papers, ensuring that essential topics, such as data repositories, are covered. He also wants to include an initial assessment to determine if new trainees are familiar with certain topics, allowing them to skip unnecessary training if they are already knowledgeable. Another priority is to evaluate the impact more systematically. He aims to develop formal indicators of the impact of his support to improve the role he holds.

This tale serves as a model for how institutions can embed reproducibility into the daily practice of science through dedicated, hands-on support. Andrew's role not only helps researchers improve their individual work but also contributes to building a more transparent and trustworthy research culture. As research becomes increasingly complex and collaborative, such roles may become not just helpful, but essential in ensuring that scientific findings are reliable and reproducible.

A JOURNAL TALE

02

SPECIALIST JOURNAL EDITORS REPLICATE ANALYSIS AND CHECK PROTOCOL COMPLETENESS OF MANUSCRIPTS

Interviewees: *Pedro Godoy & Daniel Casali, phylogenetics editors, Journal of Vertebrate Paleontology*

The Journal of Vertebrate Paleontology, the flagship journal of the Society of Vertebrate Palaeontology, has two phylogenetics editors in its editorial team, Pedro Godoy and Daniel Casali. Their responsibility is to check the correct application of methods and reproducibility of phylogenetic analyses in manuscripts. These analyses examine the evolutionary relationships amongst organisms.

Why reproducibility matters in palaeontology

Reproducibility can be an issue for certain analyses in palaeontology. When a new species is identified through the fossil record, it needs to be placed within a phylogenetic tree describing its relationship to other species. This is usually done by starting with the most recent publication that includes the phylogenetic tree of the relevant taxonomic group. The new species is inserted into the data matrix used for that previous analysis. Then the analysis is run on the new matrix to visualise the evolutionary relationships. To validate the new evolutionary tree, the previous analysis is first reproduced, then the new species is added to the data matrix, and then a new analysis is run. Palaeontologists thus usually try to reproduce the analysis from a previous study. That is common practice, but can be tricky as authors do not routinely share their data, or do not describe the analysis in sufficient detail to be reproducible.

> *With us in this role as phylogenetics editors, the journal now wants to make sure that all analyses are reproduced before they are published in the journal. Our role is voluntary and together we review about 50 manuscripts each year.*

What a phylogenetics editor does

What Pedro and Daniel do in practice is, for each accepted manuscript, make sure that all files that are needed to repeat the analysis have been shared and are correctly formatted. For example, sometimes authors share their data matrix in text format, which is not usable in the software. Or sometimes they just share the data of one species, rather than the entire matrix. It's not so much that there's a reluctance to share. It's more that authors don't think about the format the data should be in. If it's not too time-consuming, they also rerun the analysis and check the results to see if everything is correct. The journal now also asks for the data matrix to be deposited in the MorphoBank repository. Then the data matrix can be accessed easily without needing to access the paper.

Their role is complementary to peer review. Just like staff editors would check grammatical and formatting issues of manuscripts and keep asking for changes from the authors until everything is fixed, they do the same. They keep communicating with the authors, asking them to fix things until they are satisfied that the results in the paper can be reproduced.

Writer: *Veerle Van den Eynden*

And their requirements are binding. The journal has a guideline on its website explaining the requirements for phylogenetic analysis that every author should read before submitting their manuscript. After their review, they also point authors to a webpage that shows that everything required is indeed listed there. They have the power to not let a paper get through to publishing until the requirements are fulfilled.

Highlights

> *The most gratifying thing is that we can contribute good practices that we ourselves have used in our papers for so long but that may be lacking in other papers.*

Reproducibility is key in science. If analyses cannot be reproduced, there's no point publishing it. Often in the past, we downloaded a matrix from another author, tried to reproduce the analysis and didn't get the same results. That was not uncommon and frustrating. We can now share this perspective with other authors. Plus, we also learn a lot in the process and therefore improve as researchers ourselves.

Challenges

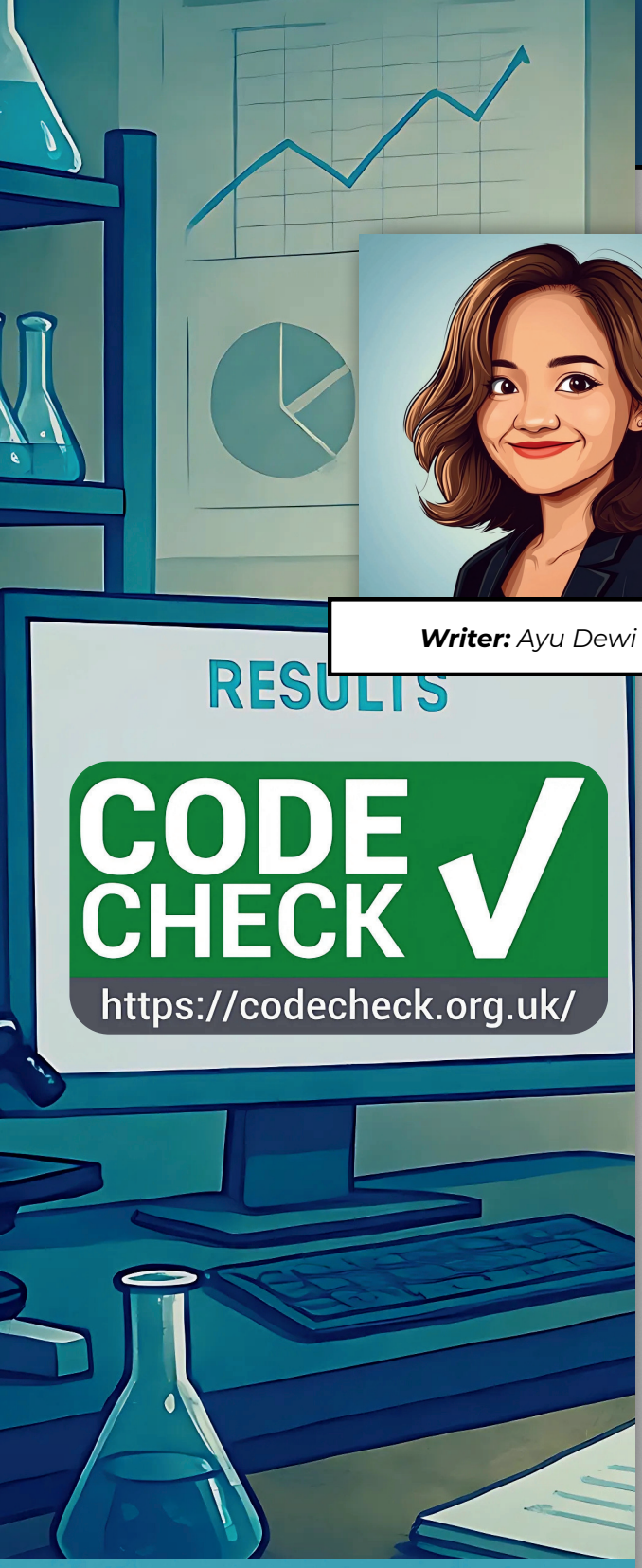
The challenge is that ensuring that all phylogenetic analyses are reproducible is a longer process than expected. It's also an educational process. Some researchers really don't think much about reproducibility. The training they received in their science education is lacking. So, it takes longer but is worthwhile. Sometimes the same researchers keep making the same mistakes. Where things seem to go wrong is when researchers describe new fossil species. They do lengthy descriptions of the anatomy and leave the phylogenetic analysis to the end, as one last thing to do, rather than a main result.

Potential for wider applicability

There is certainly potential for other journals that publish palaeontology or phylogenetic research to implement something similar. This kind of research is also published in journals with a wider scope. Some of those may have data editors, although they focus more on checking statistics. They may check manuscripts for the reproducibility of code, but not specifically for phylogenetic analysis. Phylogenetic analysis has been overlooked in the area of reproducibility. So, the Journal of Vertebrate Palaeontology is a pioneer in this respect. But hopefully others will follow, especially since future analyses frequently build on previous ones. If current analyses are not reproducible, this can cause a lot of problems in the future. Some researchers are organising short courses on a voluntary basis on this topic. That is good, because young researchers usually learn these analysis techniques from their supervisors. It's not commonly included in the methods teaching courses.

Expertise, patience and attention are crucial

The skills needed to be a good phylogenetic editor are, first of all, a good knowledge of the software used for the analyses. They each had different expertise coming to this role, so they also learned from each other. A lot of patience is needed, because it can take some time to explain to an author everything that is needed for the analysis to be reproducible. Pedro and Daniel often must repeat the same comments to many different authors. And attention to detail is needed to make sure not to skip important comments.



Writer: *Ayu Dewi*

```
x <- c(1, 2, 3, 4, 5)
y <- c(2, 4, 6, 8, 10)
plot(x, y)
model <- lm(y ~ x)
abline(model)
```

THE RESEARCH COMMUNITY TALE

03

ENHANCING REPRODUCIBILITY WITH CODECHECK AT AMSTERDAM UMC

Interviewee: *Sam Langton, Research Software Consultant, Amsterdam UMC*

Computational reproducibility, which ensures that the same data and analysis produce the same results, remains a major challenge in research across disciplines. CODECHECK[2] is a community initiative to solve this challenge through independent verification of computational workflows in scientific publications. This tale shows how CODECHECK is implemented at Amsterdam UMC (AUMC) to enhance reproducibility, address technical barriers and foster a culture of transparent research.

What is CODECHECK?

CODECHECK allows independent code checkers to rerun code, assess its reproducibility, and give constructive feedback. This process not only helps authors refine their documentation but also ensures that published findings are computationally sound.

CODECHECK pairs computational reviewers with subject-matter experts. It offers a dual-layer validation of research outputs such as tables, graphs and statistical analyses, to accurately reflect the results generated by the code. It is a free and open tool that operates independently of journals, allowing researchers to verify their research findings before the manuscript is submitted to a journal.

A successful code check earns a certificate as public endorsement of the study's reproducibility. It boosts credibility and showcases a commitment to open science.

CODECHECK's broader application

CODECHECK is not limited to one field. While it has gained traction in neuroscience and geographic information systems, its reach is expanding. Fields such as medical sciences, social sciences and digital humanities are increasingly adopting the initiative.

Researchers around the world can incorporate CODECHECK into their research processes in a way that suits their needs. The system is designed to be flexible, allowing different types of studies to benefit from its reproducibility checks. Its adaptability makes it a standout choice for institutions looking to elevate their reproducibility frameworks.

Implementing CODECHECK at Amsterdam UMC

AUMC joined CODECHECK as part of a project focused on improving reproducibility by refining the management of research software and data. The initiative was built on previous successes, like the neuroscience reproducibility hackathon and collaborations with the Dutch Reproducibility Network. Unlike traditional audits, CODECHECK at AUMC was designed to identify practical barriers to reproducibility, such as IT infrastructure constraints and documentation gaps.

The process

CODECHECK at AUMC follows a structured approach.

- Researchers submit their code and data for verification
- An independent code checker reruns the code
- The code checker documents the process, identifies any discrepancies, and provides detailed feedback via a standard template
- If the research meets reproducibility standards, a CODECHECK certificate is issued, which can be published on platforms like Zenodo.

Training and support

Anyone can become a code checker after appropriate training. Founders Daniel Nüst and Stephen Eglen offer hands-on guidance, including one-on-one mentoring, to help researchers build the skills they need. While CODECHECK can serve as research support within institutions, it also invites PhD and master's students to conduct checks for their peers, fostering a collaborative learning environment.

At AUMC, CODECHECK is an informal process, but is expected to integrate with existing research support services. In future, researchers may access it through a ticketing system or as part of data management planning requirements. Embedding reproducibility checks into the research ecosystem aims to make them standard practice and foster a peer-to-peer verification culture.

Challenges

Despite its benefits, implementing CODECHECK at AUMC has revealed several challenges. It needs open data and code, when much research uses confidential or proprietary data. This complicates reproducibility checks. While anonymisation and synthetic data can sometimes provide solutions, institutional policies must support open data practices to facilitate broader adoption.

Running computational workflows requires access to appropriate IT infrastructure, including compatible software environments and sufficient computational resources.

Encouraging researchers to prioritise reproducibility remains an ongoing effort. Institutional incentives could counterbalance the perceived burden.

Future directions

For CODECHECK to have an impact, efforts must go beyond code checking alone. That means fostering a culture where researchers, especially early career researchers, verify each other's work, even beyond formal CODECHECK reviews. Integrating CODECHECK into institutional policies, such as requiring reproducibility checks for grants or research outputs, could further encourage adoption and ensure long-term viability.

Reflections

Reproducible research relies on two key factors: open data and open code, along with clear and thorough documentation. It's like giving someone a recipe with all the ingredients but leaving out the step-by-step instructions; without those details, they won't be able to recreate the dish exactly as you made it. The same applies to scientific research. Amsterdam UMC's adoption of CODECHECK is a step toward ensuring that studies are not just credible in theory but also verifiable in practice. By embracing initiatives like CODECHECK, we move closer to a future where transparency and reproducibility are at the heart of scientific discovery.

[2] https://codecheck.org.uk/

Writer: Stefan Wouters

A FUNDER TALE

04

DUTCH RESEARCH COUNCIL NWO FUNDS REPLICATION STUDIES

Crumbling foundation

When research findings can be independently verified and replicated, science moves forward with confidence. So while replication is crucial, it may be rather neglected compared to novelties and innovation. Repeating your own or someone else's research, however important, is not quite as prestigious as new research, and may be hard to get funding for. Generally, journals rarely publish replication research, grants rarely fund it, and researchers rarely pursue it.

In 2015, a landmark study shocked the scientific world[3]. Researchers attempting to replicate 100 psychology experiments showed that only 36% produced consistent results. This replication crisis exposed the troubling truth that even high-profile findings often crumble under scrutiny. Assuming there were no fraudulent publications, the scientists behind these 100 studies believed they were performing valuable experiments and publishing important knowledge. Why then this finding? Did they lose sight that replication underlies the reliability of science? The waning awareness of the importance of replication, in combination with bias and less-than-ideal incentives, led to a growth of fragile findings, wasting resources and eroding public trust. To turn the tide, replication must receive a higher regard.

Funding replication research

In 2016, a year after this landmark study, the Dutch Research Council (NWO) responded by launching the world's first dedicated replication funding program[4]. Their goal: to restore trust in science by making replication studies mainstream. At first, their call focused on biomedical research, where replication is, at least theoretically, relatively easy to achieve. In later calls, realising a shift is needed across all of science, the discipline's focus opened further. Between 2016 and 2020, three rounds, each with a budget of €3 million, funded 24 projects. Most projects led to multiple publications, with several successful replications.

In the current round, researchers from all disciplines that are associated with Dutch knowledge institutes, including universities of applied sciences, can apply. Full reproduction studies, where original data are re-analysed using the original method, can receive up to €100.000 of funding. Double that amount can be requested for replication studies. This can be exact replication, where a study is repeated with new data, or less exact replication, where methods can be minimally adapted, for instance, when targeting a different demographic. With a budget of €5.2 million, NWO signals that verifying past work is as vital as new discoveries. This legitimisation may be part of the reason for the success of this program: might (or in this case, money) makes right.

Since transparency is a necessity for reproducibility, funded researchers have to preregister their methods, share their data openly and apply the FAIR data principles. Practices like these not only aid future reproducibility. Normalisation also further increases awareness and helps to revive replication as a standard part of science.

Making way

Although being the first replication program, fortunately, it is no longer the only one. NWO's initiative inspired similar efforts in Canada and Germany. In the Netherlands itself, the Royal Netherlands Academy of Arts and Sciences (KNAW) noted in a 2018 evaluation that, although replications only make up a small portion of publications, this amount is increasing. The support of NWO is a vital tool enabling this growth.

Besides the replication program, NWO also incentivises more Open Science initiatives, digital competency, and fair rewards and recognition, for instance, through their dedicated daughter organisation, Open Science NL. In the future, replication studies will become even more prevalent as the plan is to integrate them into mainstream funding programs. Time will tell what ratio of replication to new research bears the most fruit. For now, it is safe to say we are not yet close.

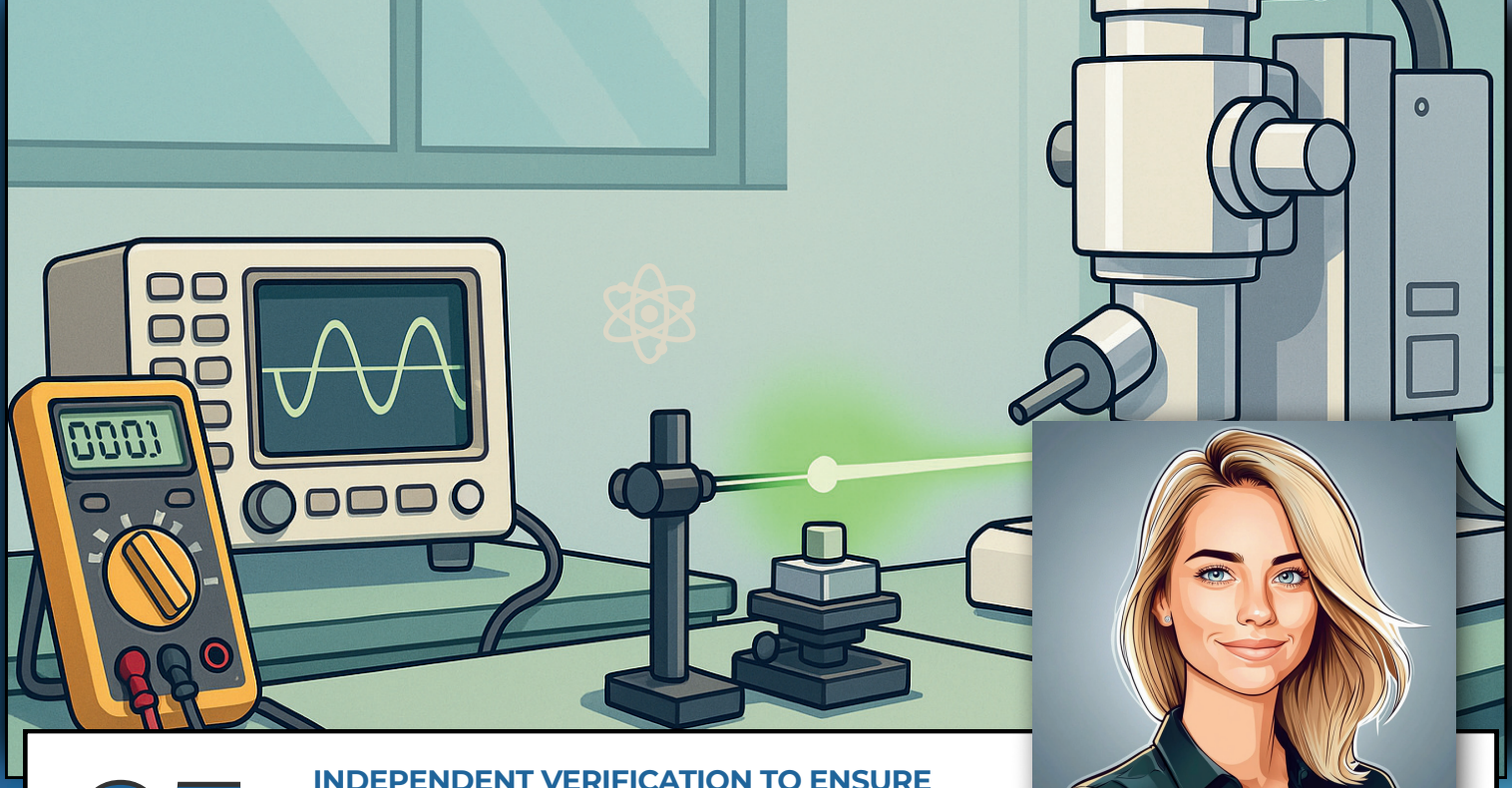
> *By valuing replication as highly as innovation, funders encourage researchers to verify and validate existing findings and ensure that scientific knowledge is built on solid foundations.*

The replication study program can be used as a model by other research councils and funding agencies. Directly funding replication enhances the reliability of scientific research, all the while improving transparency and robustness. By valuing replication as highly as innovation, funders encourage researchers to verify and validate existing findings and ensure that scientific knowledge is built on solid foundations. In an era of AI-generated content and viral misinformation, replication studies may be even more important than ever before.

[3] *Open Science Collaboration (2015). Estimating the reproducibility of psychological science. Science 349: aac4716. Doi 10.1126/science.aac4716*
[4] *Open Science NL Replication Studies programme: https://www.openscience.nl/en/researchprogrammes/replication-studies*

Supporting reding material - dig deeper, discover more!

- *Estimating the reproducibility of psychological science | Science*
- *Replication Studies | Open Science NL*
- *Call for Proposals - Replication Studies*
- *Open Science NL Work programme 2024-2025*
- *Replication studies - Improving reproducibility in the empirical sciences*



05

INDEPENDENT VERIFICATION TO ENSURE PUBLISHED DATA ARE CORRECT

Interviewee: *Marc Assael, Editor-in-Chief, International Journal of Thermophysics*

Writer: Magdalena Kozula

A JOURNAL TALE

In today's world, data can appear on a screen in seconds, polished, formatted and seemingly reliable. But as Marc Assael points out, "you can't know if they are correct or not."

For researchers working with experimental measurements, especially in highly technical fields like thermophysics, this uncertainty is not just a minor detail. It is a fundamental challenge. Some errors might be obvious, but many are not. That is why rigorous validation is so important.

> *An editor-in-chief cannot check the numbers, and reviewers usually can't either.*

A collaborative model for verifying data

Marc Assael serves as editor-in-chief of the International Journal of Thermophysics, where he oversees an innovative partnership with the Thermodynamics Research Center (TRC) at the National Institute of Standards and Technology (NIST) in Boulder, Colorado. The Journal submits every manuscript in the scope of cooperation (properties of pure compounds and mixtures) to TRC for independent checks of accuracy, plausibility and internal consistency before publication. "They check all our numbers before we publish them. And this is not a simple validation. It's a rigorous consistency analysis grounded in decades of international scientific data."

The partnership is aimed at experimental research on thermophysical properties of matter, such as viscosity, thermal conductivity and other key properties of fluids and mixtures. Most manuscripts submitted to the journal include entirely new measurements. If those measurements are wrong, they risk becoming incorrect reference points for future research, leading to a chain reaction of flawed science.

This cooperation takes a direct form, whereby a TRC expert serves as one of the peer reviewers. First, they check that the data are reported in a numeric form and well defined. If presented data requires deeper investigation, it often exceeds the capacity of verification that the editor-in-chief or individual reviewers alone have. This is where TRC comes in. Its role should not be confused with simple validation. TRC's experts compare new thermophysical property data against an extensive historical database, derive the reference correlation and validate uncertainty estimates. This ensures the methodological reproducibility of submitted manuscripts. It is especially crucial in light of common misunderstandings about measurement uncertainties.

> *A lot of people don't understand the word uncertainty. They confuse it with how many digits your instrument displays. This is the precision of electronics and has nothing to do with actual measurement. So, they usually wrongly call it uncertainties.*

Strengthening research integrity

Beyond numerical accuracy, this cooperation also addresses additional critical quality factors. Cooperation with TRC helps to identify improper reporting of experimental conditions and attempts to artificially inflate citation counts through so-called “fishing” papers – publications produced primarily for self-citation purposes rather than for their scientific contribution.

Marc notes that the process makes his job much easier and significantly improves the quality and credibility of the journal. It is not just about catching errors. It is much more about setting a gold standard for how experimental science should be reported and checked before publication, so that other researchers and readers can trust these numbers. Other journals that publish thermophysical property data follow this same procedure. Similar partnerships established by other journals could enhance reproducibility across various scientific disciplines.

The future of review checks and cooperation

This specific case is based on voluntary cooperation. TRC experts perform data checks as a peer review service. This makes the collaboration vulnerable to different challenges that academia encounters, and potentially difficult to replicate in other settings. As Marc puts it, “If this stops, I would be very worried about what I’m publishing.”

What is the future of such partnerships? Marc believes that scientific publishers should consider formalising and funding this kind of service. As artificial intelligence (AI) tools become more common, the risk of fabricated or unverified data sneaking into the literature increases. Tools to detect AI-generated text are evolving, but for numbers, human expertise backed by trusted data cannot be substituted. The key is infrastructure and the willingness of publishers, universities and policymakers to invest in both.

This tale provides an essential perspective on reproducibility by highlighting the importance of a structural approach in the academic environment. Integrating these kinds of checks of manuscripts into the publishing workflow is a sustainable solution that promotes a culture where accurate and transparent research data becomes the norm rather than the exception. Ultimately, independent verification is the most crucial factor for improving reproducibility. “Numbers must be correct,” Marc insists. Without rigorous validation, science risks spreading inaccuracies, undermining trust, and weakening research quality.

06

ROLE MODELLING: SUPERVISORS PAVE THE WAY IN OPEN SCIENCE

A SUPERVISOR TALE

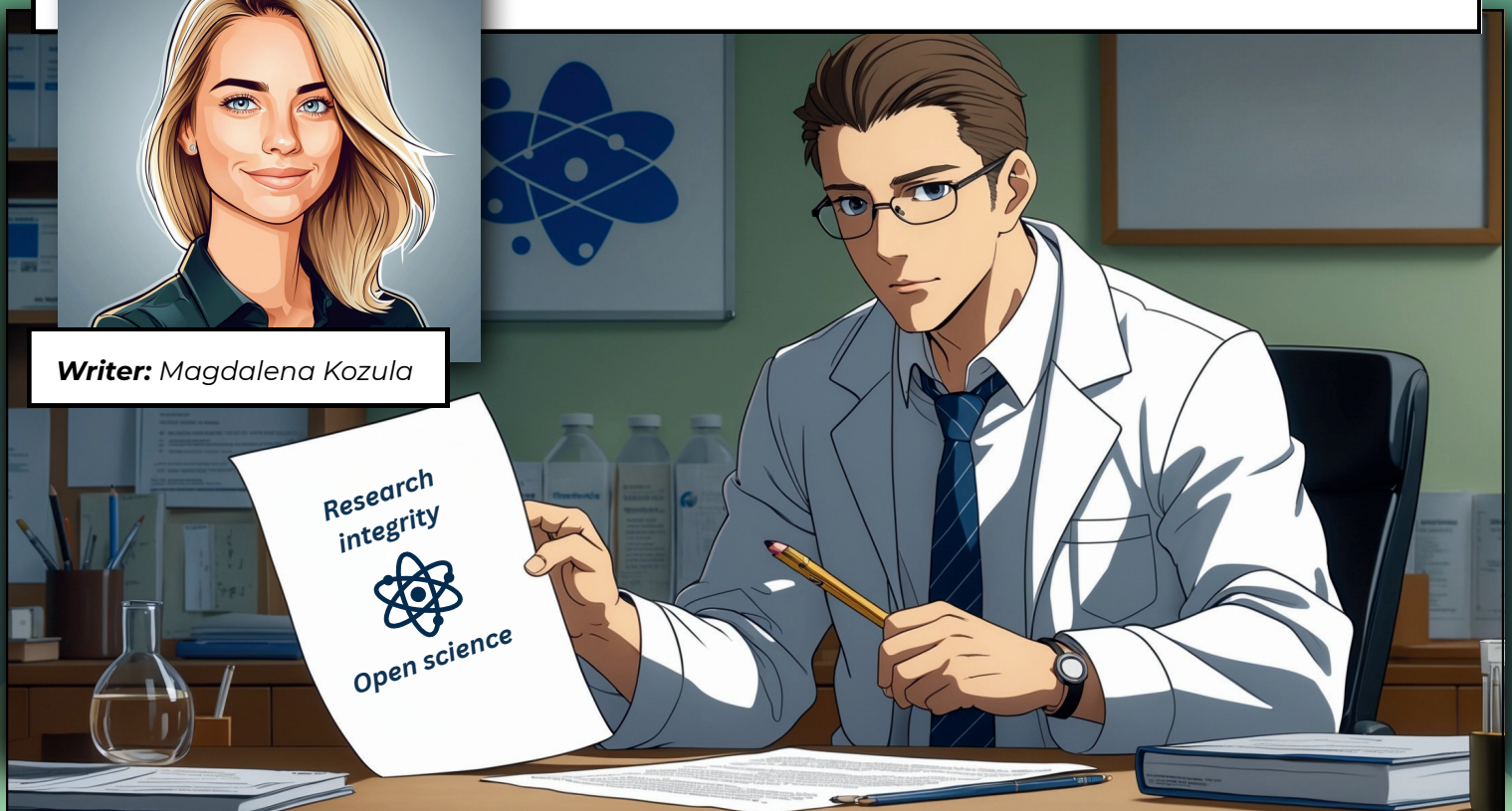
The practice of sharing data openly and publishing research in open access formats is widely recognised as crucial for reproducible research. It increases transparency and enables others to track the process behind research and verify its findings. Several initiatives have been undertaken by higher education institutions, journals and funders to encourage and embed openness and transparency in research practices. While many of these initiatives rely on formal training or are implemented through journals and funders' policies, one recent investigation concentrated on a powerful mechanism of social learning in academia: role modelling by senior researchers.

The study specifically targeted biomedicine, a field where reproducibility is especially important. When research findings in this area are not reproducible, it can have a direct impact on patient care and the development of medical treatments. Health or even lives can be at stake when findings cannot be replicated. In this context, open access is also seen as a way to facilitate timely access to research, potentially accelerating scientific discovery and its translation into clinical practice.

Open Science practices of supervisor and PhD papers

Tamarinde Haven, Susan Abunijela and Nicole Hildebrand examined 2,062 empirical publications produced by 211 pairs of PhD candidates and their supervisors at four Dutch University Medical centres[5]. Of these, 651 papers were first-authored by the PhD candidates, while 1,411 were first- or last-authored by the supervisors.

As the researchers aimed to understand whether supervisors' practices around data sharing and open access publishing influenced the behaviour of their PhD students, they compared a candidate's behaviour with a separate snapshot of their supervisor's typical practice. To learn how often each publication was openly accessible, they used Unpaywall, an open database that provides access to more than 23 million open-access research publications and resources. They also employed Oddpub, a text-mining algorithm, to detect statements about shared data within each paper. Any articles where Oddpub spotted a potential data sharing language were then inspected manually to confirm actual data sharing. By collecting data from multiple institutions, the researchers ensured relevance across various biomedical sub-disciplines.



Writer: *Magdalena Kozula*

What do the numbers reveal?

The findings showed that supervisors play an important role in shaping their PhD candidates' engagement with open science practices, particularly when it comes to data sharing. Early career researchers were significantly more likely to share their data if their supervisors did, with the effect strengthening after accounting for false positives. After manually removing cases where a data availability statement was present in the paper, but no real data were shared, the odds rose from 2.2 to 4.6 times. A similar trend was observed for open access publishing. Here, PhDs with open-access-active supervisors were nearly twice as likely to follow suit. However, the latter sheds light on some challenges involved in studying such behaviours. This effect lost statistical significance after adjusting for institutional differences. suggesting that the relationship observed may stem from institutional clustering rather than a direct effect of supervisory behaviour on PhDs' practices. Meanwhile, privacy concerns related to patient-derived data pose their own hurdles to sharing data openly.

A key limitation of this study is that the applied design did not make it possible to distinguish who might have initiated open science practices, as motivated PhD students may also favourably influence their supervisors rather than the other way around. Still, it did highlight the strong correlation between supervisors and PhD candidates' behaviours.

Mentor influence for open science

Why does investigating this perspective matter? While these findings point to both the importance of supportive mentorship and the structural factors that shape open science adoption, academia has long functioned as a space for social learning, where early career researchers develop their practices by following senior scholars. Role modelling has been a foundational part of academic training for centuries and is presumed to be a crucial component of responsible supervision.

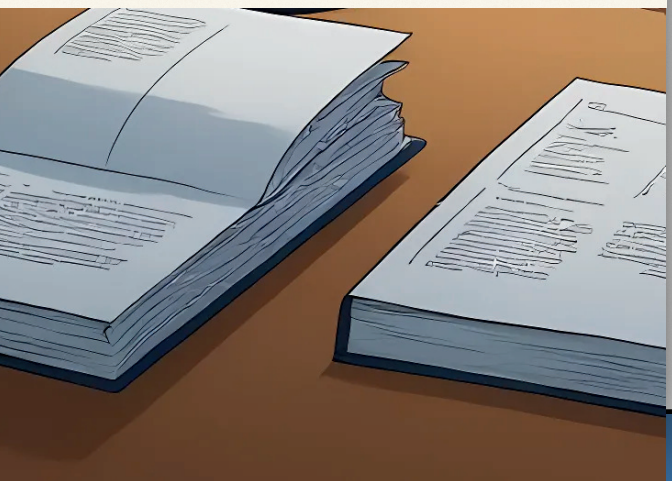
Looking forward, the authors suggest several promising directions for future research. Longitudinal studies and targeted interventions could help clarify cause-and-effect relationships, while broader international studies may help to understand how applicable these findings are across various fields and contexts. They also recommend examining how additional factors, such as institutional training programs or financial incentives, might further encourage open science practices.

This tale highlights a straightforward yet powerful message: when supervisors actively engage with open science, they're not only supporting open science within their own teams but also contributing to a broader cultural shift in research.

Perhaps the most effective way to strengthen reproducibility could be through example. Since academia at its core is a community, exploring this direction seems like an organic way to create a research environment based on openness and transparency.

[5]*Haven, T. L., Abunijela, S., & Hildebrand, N. (2023). Biomedical supervisors' role modeling of open science practices. eLife, 12, e83484.* https://doi.org/10.7554/eLife.83484

Writer: *Paula Muñoz*



A PEER REVIEW TALE

HOW PEER REVIEWERS CAN ENHANCE REPRODUCIBILITY 07

Interviewees:
Gaëlle Le Pavic, United Nations University; David Smailes, Northumbria University; & Czifrus Szabolcs, Budapest University of Technology and Economics

Reproducibility is one of the core values of science. Yet achieving it remains a persistent challenge. Without reproducibility, we cannot fully trust or build upon published research. This undermines scientific progress.

Surprisingly, one of the most common obstacles to reproducibility is cognitive rather than technical. When researchers write up their own studies and experiments, they often assume that certain steps or decisions in the methodology are self-evident. Thereby, they unintentionally omit details essential for replication. This is where an external perspective becomes essential. A fresh pair of eyes can identify gaps in clarity and logic that the authors may overlook.

Fortunately, science already has a mechanism for external evaluation: peer review. Where peer review traditionally focuses on assessing the quality and rigour of research, it also holds untapped potential to enhance reproducibility. The challenge is that most reviewers are not trained to identify potential flaws for reproducibility in a manuscript. And there is no standardised process for it.

This issue was highlighted by Gaëlle Le Pavic of United Nations University, David Smailes of Northumbria University and Czifrus Szabolcs of Budapest University of Technology and Economics. In their experience, the quality of peer review can vary. They appreciate constructive reviews where the peer reviewer gives specific comments, avoids bias and picks up on inaccuracies, discrepancies or problems..

They propose five practical recommendations to incorporate reproducibility checks into the peer review process.

Check the use of reporting guidelines

Peer reviewers can check whether a manuscript adheres to relevant reporting guidelines, such as CONSORT (Consolidated Standards of Reporting Trials) for clinical trials or PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) for systematic reviews. Reporting guidelines outline essential components that should be included when describing methodology and findings in research reports. That ensures that critical information is consistently presented. This then allows other researchers to repeat the same methodology, thus facilitating reproducibility.

Check that the methodology is clear

A clear and detailed methodology description is crucial for reproducibility. When the methods used in a study are explicitly described, with all relevant parameters listed, other researchers can replicate the study's procedures accurately. This includes specifying experimental designs, data collection techniques, and analytical methods. Ambiguities in the methodology can lead to different interpretations and outcomes, undermining reproducibility. If a study was preregistered, then check during peer review that the methodology and analysis in the manuscript match the preregistration.

Check that data, code and materials are openly available

Providing access to all relevant data, materials, and resources is essential for reproducibility. This means sharing datasets, code and supplementary materials. When reviewers ensure that authors make this information available or indicate where it can be accessed, it allows other researchers to verify results and conduct follow-up studies. For qualitative research, where underlying data may need to be restricted for privacy reasons, sharing the analytical coding tree with coded extracts can make the analysis process transparent and facilitate reproducibility of findings.

Check transparency about limitations

Acknowledging the limitations of a study is vital for reproducibility. When authors are transparent about potential biases, confounding variables or constraints in their research, it helps other researchers understand the context and applicability of the findings.

This transparency allows for better informed replications and helps identify areas where results may not be generalizable. In qualitative research, a positionality statement is an important part of transparency, since a researcher's positionality influences data collection and analysis.

Give constructive and focused feedback

When suggestions and feedback are overly extensive, too generic or ambiguous, it becomes difficult to comprehend what to change or improve and can be counterproductive. The interviewees agree that it is extremely important for the review report to be specific, thoroughly explained and constructive. To do this, they recommend providing comments with suggestions for improvement directly in the document. This allows a concrete understanding of what to improve and directs action to do it effectively. Their experience is also that special issue guest editors often provide very constructive input that benefits the quality of papers.

> *I think open peer review is a way to show the quality of the review. And peer review is probably the main lever to pull to improve reproducibility in science.*

Finally, the quality of peer review can be improved with support and training. Institutional support can help here by developing review guidelines with clear indicators to standardise the peer review process. And peer review training for early career researchers, by publishers or institutions, can equip them with the skills and knowledge needed to perform thorough and effective reviews.

By taking these practical considerations into account and through adequate support and training, peer review can be an effective tool for promoting reproducibility.

Writer: Stefan Wouters

A RESEARCH INSTITUTION TALE

08 AIR YOUR DIRTY LAUNDRY

Interviewee: *Wiesje van der Flier, Alzheimer Center Amsterdam*

Have you ever made a mistake in your research? Or have you encountered an ethical dilemma? If your answer is no, you are either the luckiest scientist alive, or — more likely — you are lying. Most scientists believe their work must be perfect, untarnished and infallible. While science may theoretically approach this perfection, research is performed by imperfect humans in an imperfect world. The standard we set for ourselves is practically unattainable. Nonetheless, many scientists hold themselves to this standard, making it taboo to speak up about mistakes. This closed culture hampers reproducibility and may eventually even drive motivated scientists away from research. The culture surrounding mistakes must change.

Not so HOT

Researchers often fear that admitting or speaking up about mistakes or unethical issues leads to career repercussions, a damaged reputation or loss of funding. This fear discourages open discussion and correction of errors, as researchers may opt not to fully document or disclose all details of their methods. This hampers reproducibility.

> *Everybody makes mistakes, even senior researchers, even after twenty years of experience. We all make mistakes. What counts is how you act when they happen.*

Hiding mistakes also prevents learning from them. Without an open culture, one cannot refine techniques, nor improve the overall quality of research. This can lead to repeated errors, which in turn cost time and effort to correct. Colloquially this is known as the bullshit asymmetry principle. Honest, open and transparent (HOT) communication is essential to ensure that research findings are reliable and reproducible. The taboo on mistakes perpetuates a culture where perfection is expected, rather than a culture of continuous learning and improvement.

Bare your butt

In 2015, due to concerns about research integrity and sloppiness, Wiesje van der Flier, the scientific director of the Alzheimer Center Amsterdam, initiated "*met de billen bloot*" sessions. This literally translates to "bare your butt", but can be more aptly translated to "air your dirty laundry". In these annual sessions, attended by the entire staff, researchers are invited to openly discuss mistakes, ethical dilemmas and methodological flaws in their work. Initially, senior researchers would kick off the sessions, leading by example to show that everyone can make mistakes. Then everyone can join in to share their experience. The goal here is not to discuss all issues in a single day, but to make sure that researchers are not afraid or embarrassed to discuss issues at any time. While the sessions mainly promote research integrity and open communication, they foster a constructive culture where transparency becomes a tool to improve research quality and reproducibility. As such, the sessions reframe mistakes as opportunities for collective improvement.

In 2018, the sessions expanded to have a junior-only meeting, offering a safe space to openly discuss issues concerning seniors or supervision. Creating a safe environment is crucial for open conversation. By prioritising social safety over hierarchy and perfectionism, researchers feel more comfortable and heard. Indeed, researchers are overwhelmingly positive about the "*met de billen bloot*" initiative. They also seem more willing to share their concerns during the rest of the year. While success hinges on pre-existing openness, to the point that less collaborative teams may struggle to adopt similar sessions, the implementation is simple and cheap. Any research institution can replicate this initiative.

Wider applicability

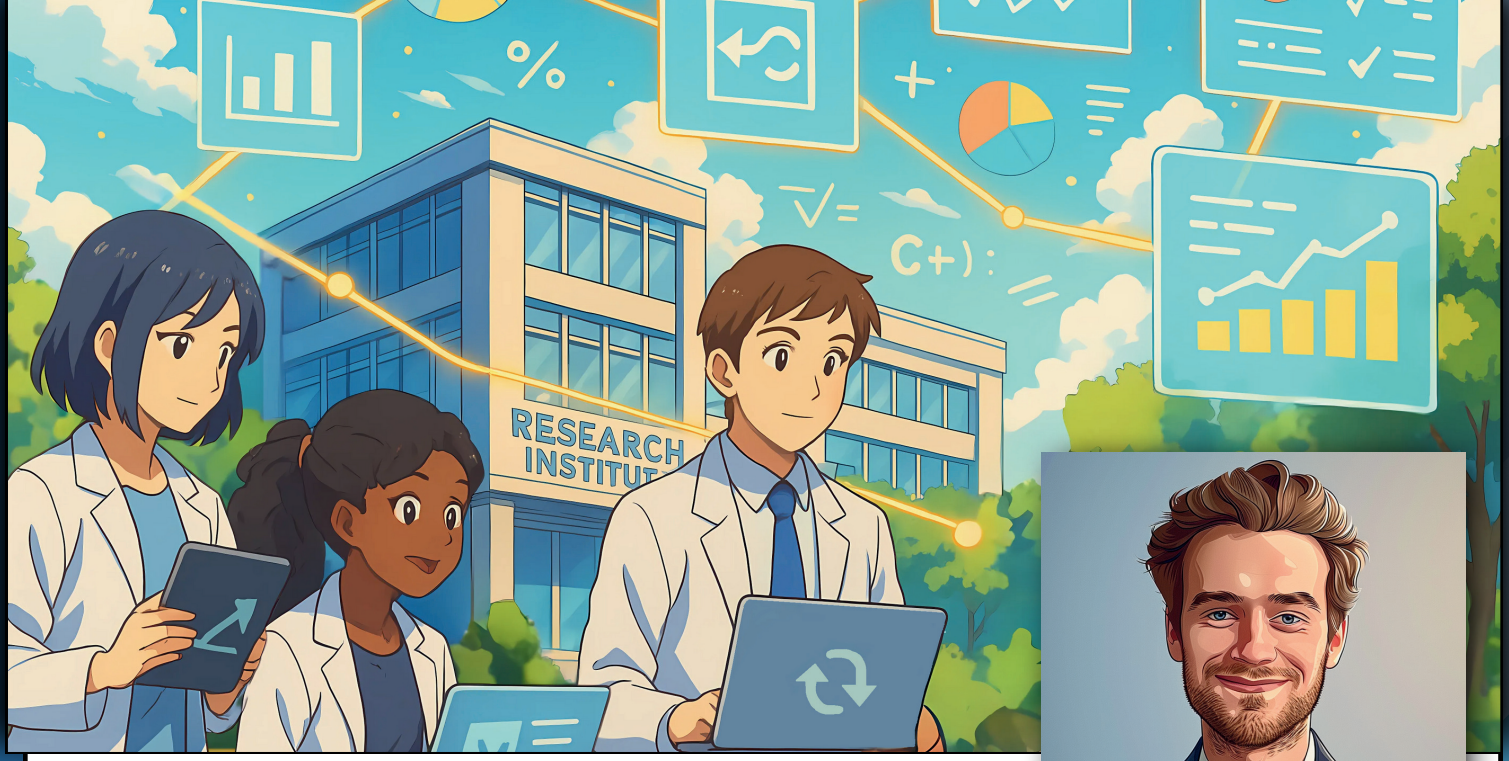
What is the recipe for success? Three factors are vital. First, senior researchers should lead by example, showing their own vulnerability. This way, they can dismantle hierarchies and create a socially safe space for sharing concerns. However, having a safe space should not just depend on the willingness or capability of seniors. This brings us to the second point. Separate meetings for the full team and for juniors only can ensure that everyone can freely and comfortably speak their mind. The third and final factor is participation. For the Alzheimer Center, the annual sessions were obligatory to make sure that nobody is left out of the culture change, and an open community can be built. If participation cannot be compulsory, it should at least be incentivised.

The "*met de billen bloot*" sessions demonstrate that breaking the taboo on mistakes is necessary to enhance reproducibility. By fostering a culture of transparency and accountability, researchers can ensure that their work is rigorous, reliable and replicable. This approach not only improves the quality of research but also prevents research waste in trying to correct mistakes.

All in all, airing your dirty laundry can support an honest, open and transparent research culture. A research culture that prioritises social safety over perfectionism means scientists can address flaws before they compromise reproducibility and can learn from them.

Supporting reding material - dig deeper, discover more!

- Sofasessie met Wiesje van der Flier - Alzheimercentrum Amsterdam
- "Met de billen bloot" (airing your dirty laundry)
- #billenbloot #blijmetmijncollegas #research #integrity #sloppyscience... | Wiesje van der Flier



09

FIXING THE REPRODUCIBILITY CRISIS THROUGH REPLICATION: THE INSTITUTE FOR REPLICATION

Writer: Constant Vinatier

Interviewee: *Abel Brodeur, Chair of the Institute for Replication*

A COMMUNITY TALE

In 2022, economist Abel Brodeur established the Institute for Replication (I4R)[6], an independent initiative committed to addressing the challenge of research reproducibility. Having encountered problems such as p-hacking and publication bias early in his career, Brodeur sought to develop concrete solutions. After securing a permanent academic position, he took a year to focus on establishing I4R, with the goal of systematically reproducing and replicating published studies across multiple disciplines.

> *I have always been interested in questions of research transparency: do researchers engage in p-hacking? Do they share their data? Do they follow good research practices?*

Initially focused on economics, I4R has since expanded into political science, psychology and environmental studies, and is now preparing to enter public health research. The objective is simple yet ambitious: to evaluate the robustness of research findings by independently reproducing studies using publicly available data and methodologies.

A collaborative and scalable model for replication

I4R was founded on a simple yet fundamental question. How do we take someone unfamiliar with reproducibility and bring them into the world of replication? At the heart of the I4R approach lies a highly collaborative model. The institute mobilises researchers from different fields through two primary methods: targeted replications and replication games.

Targeted replications are done when a new study is published with accessible data. I4R then seeks experts willing to independently reproduce the findings.

Replication games function like hackathons, where researchers, in teams of four or five, choose a study from a predefined list within their field. Over the course of a day, they work on reproducing the results by re-running the statistical analyses, checking for inconsistencies and assessing the robustness of the study. At the end, they submit a detailed replication report, which is sent to the original authors. This initiates a dialogue, where authors can respond, before the reproduction findings are publicly shared on I4R's website with the response from the authors.

This model has gained international traction, with 29 replication games held in 15 countries last year, involving 1,300 researchers ranging from PhD students to senior researchers. The initiative also fosters partnerships with major academic journals. For example, I4R collaborates with Nature Human Behaviour and Psychological Science to reproduce and replicate research articles published in these journals.

Crucially, I4R maintains strict independence to ensure that its reproductions and replications remain impartial. The institute publishes its findings — positive or negative — without influence from authors or institutions.

A cultural shift in research practice

Expanding replication efforts to new fields comes with significant challenges. While some disciplines, such as economics and political science, have embraced open data and code sharing, others remain resistant to transparency. Limited access to raw data and statistical code can obstruct replication efforts, highlighting broader issues in scientific publishing.

Participation in replication studies has an unexpected benefit here, as it directly improves research practices. By attempting to replicate a study, researchers experience firsthand the difficulties of deciphering another scientist's methodology. This leads them to adopt better data management and documentation practices in their own work.

The response from the scientific community has been largely positive. Unlike traditional reproducibility research, which often focuses on exposing irreproducible studies, I4R's findings reveal the more balanced reality that many studies are reproducible and robust. In these cases, authors often welcome the validation and appreciate having their work independently confirmed.

Ambition and future directions

I4R is expanding its efforts by addressing major societal challenges. Rather than focusing on specific journals or fields, they now aim to reproduce all studies related to a given research question. One of their latest initiatives is a large-scale program dedicated to reproducing studies on deforestation. By systematically verifying research on this critical issue, the Institute aims to strengthen the evidence base for policy making and ensure that decisions regarding environmental protection are grounded in robust, reproducible science.

I4R represents more than just a replication initiative. It is an attempt to reshape scientific norms. The long-term vision is to make reproduction an integral part of the research ecosystem, where independent verification is the norm rather than an exception.

> *Launching the institute was easy. The real challenge is achieving my ambition, that in 5 to 10 years we will have fundamentally changed the standards of scientific research.*

By advocating for transparency, rigorous verification and a shift in research culture, I4R is helping to strengthen trust in science and support political decision making with robust and reliable information. In a world where misinformation spreads rapidly, ensuring the credibility of scientific research is more critical than ever.

[6] https://i4replication.org

A COMMUNITY TALE

10 MAKING YOUR OWN TALE OF REPRODUCIBILITY WITH RESOURCES AND TRAINING

After reading the previous nine Tales of Reproducibility, we want to use the last tale to highlight general advice and resources that can help you create your own Tale of Reproducibility! We call attention to useful platforms to enhance your research practice, highlight open resources for self-guided learning, and offer final advice on how to further build your skills in these areas. While we can't highlight every resource and training opportunity that exists, we hope this general advice and direction can enable you to seek out the tools that work best for your own research practice.

It is important to remember that making your research reproducible is not an innate skill. Many practices that improve the transparency and interpretability of your research, and therefore potentially make it more reproducible to others, are generally not a standard part of academic training and need to be learned. While this is changing in some places, as topics like open science make their way into curricula, there are plenty of existing gaps for researchers across disciplines and career stages.

Writer: Nicholas DeVito

Where should I share?

A key aspect of reproducible research is making information about your study widely available. If other researchers don't have access to data, materials and code, their ability to understand what you did, why you did it, and how to do it themselves will be limited to the information available in the limited space of the published paper (if there even is a paper to check!). However, there are several platforms available, with a variety of features, that can enable sharing information about your research.

General-purpose repositories are designed to handle, organise and share information about research studies. The Open Science Framework[7] (OSF), run by the Centre for Open Science, allows basic storage, organisation and publishing of project resources. It supports a variety of features such as study registration, a preprint server, assigning persistent identifiers and integrations with various other services. We use the OSF to manage our entire OSIRIS project from start to finish. It acts as the main hub for organising and openly sharing detailed project information from across the consortium[8]. Other general-purpose academic repositories include Zenodo — run by CERN and OpenAIRE —, FigShare — a private company — and Dryad —a non-profit entity. This is by no means a comprehensive list, but these are all large, well-known, and commonly used destinations to store and publish your research data and materials. GitHub repositories or similar services may be another great resource for storing, sharing and managing project code and data and have integrated support with other repositories.

While these generalist repositories are a solution for many researchers, there are almost certainly other options available. Many research-performing organisations and funders manage their own repositories where affiliated researchers can easily deposit research data and code. In addition, domain specific repositories may also exist within your discipline. These may be designed to facilitate unique aspects of sharing data and materials in your field and directly target colleagues from within your discipline. For instance, in clinical medicine, sensitive patient data from studies is managed and shared on platforms like Vivli and the YODA Project, which helps ensure responsible data access.

It also gives you the tools you need to teach and train others. The Center for Open Science also supports a global network of COS Ambassadors who can share and provide guidance on open science topics with open resources available for anyone to use.

Your own communities may also have resources on open science and reproducibility. Expert research librarians and data stewards at university libraries have a deep knowledge of available resources to help you grow your own skills. For instance, the University of Oxford's Bodleian Libraries offer various forms of Open Scholarship Support and host an annual Forum of Open Scholarship to promote these ideas within the broader community.

> *Reproducibility Networks[12] and ReproducibiliTea journal clubs[13] are fostering growing communities interested in these topics around the globe.*

Where to start

So, how can you get started? There is no one right answer for everyone. What practices to use and how to apply them will be influenced by the resources available to you, what you're hoping to get out of them, and the specifics of your discipline. However, one thing is for sure: whatever you hope to utilise in your own research practices will take some time to learn and implement. So the sooner you start, the faster you can start making progress! Things like preregistering your study protocol, making research data FAIR and sharing code are, most fundamentally, research skills that you can improve with practice. While your first efforts might not be perfect, it's better to start your journey somewhere and do what you can than share nothing at all.

How do you know where to start? A good place would be to see if journals, funders, or other important bodies in your field have requirements, or even optional suggestions, that you can start working towards fulfilling. The Tales of Reproducibility in this collection highlight a small sample of exciting initiatives in place across disciplines. Colleagues may also have guidance from their own experiences. Or you can seek out local communities of practice that can advise on future growth and development.

> *Practising open and transparent science is a skill, not a light switch. It's something you can constantly work to improve, learn and grow at.*

Looking ahead

Trying to make your research more open and transparent, in service of making it more reproducible, can feel like a daunting task. There is a lot to learn. And support on how and where to start is not always available. That said, a little bit of work and taking advantage of the wide range of resources available can go a long way and get you started down the path.

[7] https://osf.io
[8] https://osf.io/8nyke/
[9] https://fairsharing.org
[10] https://faircookbook.elixir-europe.org
[11] https://forrt.org
[12] https://www.ukrn.org/global-networks
[13] https://reproducibilitea.org

Acknowledgments

We extend our heartfelt thanks to the researchers and stakeholders who participated in interviews and focus group discussions, on which many of these tales are based. Your willingness to engage and share openly has helped illuminate pathways toward a more robust and credible research culture. This booklet aims not only to recognize your contributions, but also to inspire others by showcasing what is possible when reproducibility becomes a shared value. Thank you for being part of this important journey.

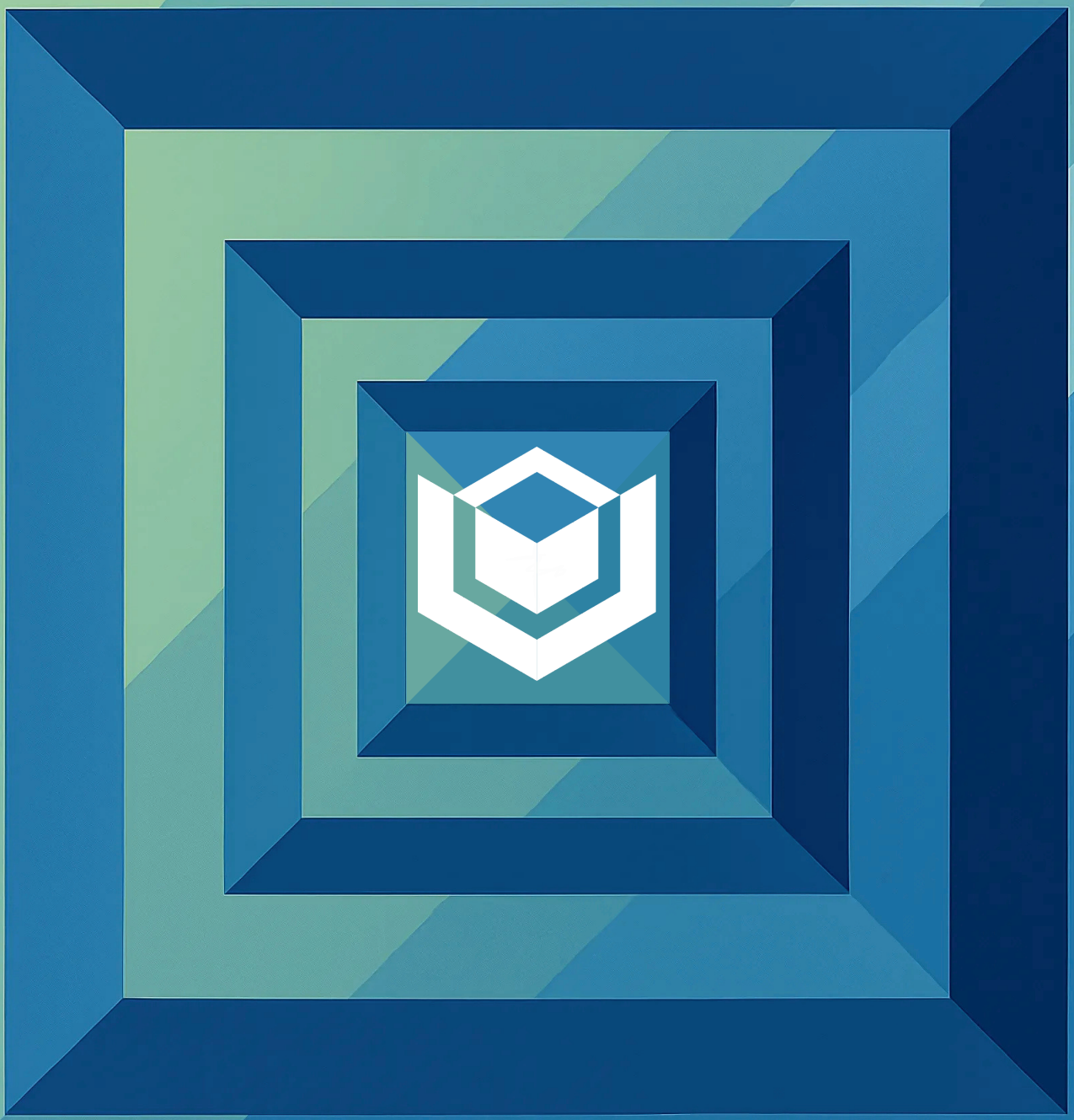
This work was supported by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101094725.

The OSIRIS Project Team

Funded by the European Union

VISIT OUR WEBSITE

osiris4r.eu



The OSIRIS project is funded by the European Union's Horizon Europe research and innovation programme under grant agreement No. 101094725. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (ERA). Neither the European Union nor the ERA can be held responsible for them.