How to Get There From Here? Barriers and Enablers on the Road Towards Reproducibility in Research

Serge P. J. M. Horbach[1], Nicki Lisa Cole[2], Simone Kopeinik[3], Barbara Leitner[4], Tony Ross-Hellauer[2], Joeri Tijdink[4]

Reproducibility of research is a hotly debated topic, including aspects like causes and consequences of low levels of reproducibility. While some research fields have led the way and introduced various reproducibility practices and procedures, the call for efforts to 'improve' reproducibility in research has not come without criticisms. The current study uses a future studies methodology to gather perceptions of developments in the research ecosystem related to reproducibility issues. It draws on input from representatives of four main stakeholder categories: scholarly publishers, funding agencies, qualitative social scientists and machine learning researchers. Particularly, it discusses the enablers and barriers that members of these stakeholder communities foresee on the road towards a research ecosystem that is more conducive to reproducibility. The study finds that enablers and barriers can be categorised into five main clusters. The factors most prominently mentioned as potentially supporting or hindering a desired future are located within research culture, including norms, values and shared definitions; and in the infrastructure required to engage in reproducibility practices, including repositories, support staff, and digital infrastructure. Three other clusters of factors put forth by participants relate to policy efforts required to incentivise reproducibility practices; training and education to empower researchers and support staff to engage in reproducibility practices; and the financial resources required to facilitate the transition towards a desired future and to specifically fund replication studies. This manuscript also identifies several tensions between enablers and barriers perceived by diverse stakeholders and concludes with recommendations for addressing these.

Keywords *reproducibility, epistemic diversity, research communities, co-creation*

[1]Institute for Science in Society, Radboud University, Nijmegen, the Netherlands

[2]Digital Transformation Design, Know Center Research GmbH, Graz, Austria

[3]Fair AI, Know Center Research GmbH, Graz, Austria

[4]Department of Ethics, Law and Humanities, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, the Netherlands

Received
March 28, 2025
Accepted
February 7, 2026
Published
March 17, 2026

Correspondence
Radboud University
serge.horbach@ru.nl

Introduction

Reproducibility, or more specifically, concerns over the lack of it, have recently received ample attention from the research community and wider public alike (Baker, 2016). The concept of reproducibility, though interpreted in various ways, generally implies the scientific community's ability to achieve results that are either identical or similar to a study's original findings by repeating the research methods or analyses (Plesser, 2018). We recognise that definitions of reproducibility and replicability vary widely. Here, we focus on reproducibility and use the term in a broad and inclusive sense of obtaining consistent results when repeating studies, experiments and analyses. Reproducibility has been a subject of increasing concern across several fields, particularly in behavioural and medical sciences, due to a perceived lack of success in replication studies (Klein et al., 2014). Researchers have suggested or identified various factors that are or might be contributing to low levels of reproducibility. These include insufficient clarity in reporting procedures, data, and analysis methods; a tendency to publish

primarily positive outcomes; and problematic research practices (Baker, 2016; Fanelli et al., 2017; Guttinger, 2020).

Like the causes of failures to reproduce previous findings, the consequences are equally debated. Some view the lack of reproducibility as a major threat to the scientific process's self-corrective nature, research efficiency, and public trust in research findings (Peels & Bouter, 2018; Sarewitz, 2016). However, the relevance of reproducibility varies greatly depending on the research context and the type of research at hand (Guttinger, 2020; Leonelli, 2018; Ulpts & Schneider, 2023). In some contexts, lower levels of reproducibility are not always deemed problematic, and reproducibility is not always considered a major concern; for example, this is often the view regarding qualitative research (Davies & Dodd, 2002; Huma & Joyce, 2022; Reischer & Cowan, 2020; Tamminen & Poucher, 2018). Factors influencing the applicability and desirability of reproducibility include the resources required for reproducibility, standardization levels within the research field, the philosophical underpinnings of the research, the subject matter being studied, as well as ethical and legal limitations on data sharing and transparency (Leonelli, 2018). In particular, there have been lively debates within the interpretative and qualitative social sciences and humanities about whether efforts to increase standards of reproducibility are relevant or even desirable and feasible for their fields of research (Bazzoli, 2022; Bennett, 2021; Drummond, 2019; Penders et al., 2019).

Improving Reproducibility

In the quest to improve levels of reproducibility, various interventions have been suggested and implemented across research communities. Most of these interventions have originated from the medical and behavioural sciences (Fidler & Wilcox, 2021), and some have gained traction in various other disciplines. One significant intervention category is open methodology, where researchers are encouraged to make their research processes transparent. This includes detailing experimental procedures, analytical methods, and decision-making processes. The rationale is that such openness allows other researchers to understand, evaluate, and replicate the studies more effectively (Fuller et al., 2014; Pownall M et al., 2023). For instance, various registries and platforms (e.g. the Open Science Framework) offer researchers tools to preregister study designs and methodologies, promoting transparency from the outset. This approach has been gaining traction, as evidenced by an increasing number of journals and funding bodies that now require or strongly recommend open methodology practices (Giofre et al., 2022; Veroniki et al., 2021). Considering the debate about different reproducibility practices in diverse research traditions, efforts have been made to tailor preregistration formats to different ways of conducting research (e.g., Haven et al., 2020). Initial studies assessing the effectiveness of preregistration in terms of fostering reproducibility show small but positive effects (Obels et al., 2020; e.g., Schäfer & Schwarz, 2019; Scheel et al., 2021; Soderberg et al., 2021). Yet, there remain concerns and questions about the validity of this intervention across diverse epistemic contexts (Branney et al., 2023; Evans et al., 2023; Haven et al., 2020; Karhulahti et al., 2022).

Another critical intervention lies in the domain of open data and reporting standards. The movement towards open data involves making research data publicly available, thus enabling other researchers to verify results, conduct new analyses, or combine datasets for more robust findings. Initiatives related to sharing of data, like the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, have been instrumental in guiding researchers towards more accessible and reusable data practices, usually through data sharing policies by funders and journals. Additionally, the adoption of stringent reporting standards, like the CONSORT guidelines for clinical trials (Betz, 2011), the ARRIVE guidelines for animal research (Percie Du Sert et al., 2020), or the PRISMA guidelines for reviews (Page et al., 2021), aspires to set minimal standards for the level of methodological detail conveyed in published research, to allow for replication and evaluation. These guidelines are often endorsed or mandated by journals. Studies find mixed but mostly positive effects of these kinds of guidelines on reporting standards (e.g., Cobo et al., 2011; Hopewell et al., 2016; van der Braak et al., 2022). While most of these initiatives, again, originate in the (bio-)medical

and behavioural sciences, some guidance also exists on reporting qualitative data and data from the humanities. However, despite these interventions, studies indicate that the prevalence of open data practices remains low and debates about the desirability and usefulness of sharing qualitative data are ongoing (Bergeat et al., 2022; Hardwicke et al., 2022; Haven et al., 2023; e.g., Open Data Barometer, 2017).

Additionally, interventions in the form of improved infrastructures and open evaluation processes are reshaping research practices. Digital infrastructures like research data repositories and collaborative platforms aim to facilitate easier data sharing and collaboration. Many such infrastructures exist; some are particularly tailored to the needs of specific research communities (e.g., Zuiderwijk et al., 2013). Open evaluation, including transparent peer review processes, has the potential to increase levels of scrutiny of research (Ross-Hellauer & Horbach, 2024). This openness in the evaluation process aims to increase accountability and improve the quality of published research. Despite these advancements, the effectiveness of these interventions in terms of reproducibility remains largely unknown (Ross-Hellauer et al., 2023).

Lastly, several stakeholders have experimented with symbolic rewards, for example in the form of tokens or badges, that can be earned for acts considered to be good research practice (Eich, 2014). Such interventions, including badges for implementing practices like data sharing, code sharing, or preregistration, aim at steering community norms and individual incentives towards higher engagement with reproducible research practices. Studies evaluating the effectiveness of these symbolic rewards to increase the likelihood of researchers engaging in the practices they are rewarded for, give mixed results. Indeed, some studies found positive effects on data sharing (Rowhani-Farid et al., 2020), but others found null effects on data sharing (Kidwell et al., 2016) and on code sharing (Rowhani-Farid & Barnett, 2018). The effects of such incentive practices on reproducibility seem to be minimal, as evidenced by a study finding low levels of reproducibility, even among studies receiving badges (Crüwell et al., 2023; Hardwicke et al., 2021).

As mentioned before, the call for efforts to 'improve' reproducibility in research has not come without criticisms. Both the initial identification of issues—quickly labelled as a 'reproducibility crisis'—and the implementation of the proposed solutions described above originated in certain disciplines, each with their own characteristics, methodological procedures, and normative frameworks (Ulpts & Schneider, 2023). Scholars from other disciplines and traditions have repeatedly questioned the desirability of establishing universal standards and employing blanket strategies in relation to reproducibility. Criticism originated most notably from the qualitative social sciences (e.g., Bennett, 2021; Pratt et al., 2020) and the humanities (De Rijcke & Penders, 2018; Holbrook et al., 2019; e.g., Leonelli, 2018). The latter critiques were voiced as a direct response to an explicit pledge for the desirability of reproducibility in the humanities by Peels and Bouter (2018). Opponents of universal standards for reproducibility argue that the appropriateness and the meaning of reproducibility starkly depend on the context and unique conditions of a study (Guttinger, 2020). These conditions may vary even within disciplines and are closely related to ways of producing and certifying knowledge (Ulpts & Schneider, 2023). Consequently, diverse research and research-supporting communities may have divergent understandings of the notion and appropriate standards of reproducibility, as well as ways of achieving it. This consequently calls for a community-driven approach towards understanding the specific enablers and barriers towards desired states of reproducibility in research.

Aims and Objectives

The study described in this article uses a future studies methodology to gather perceptions regarding developments in the research ecosystem in terms of reproducibility issues. Particularly, it draws on input from representatives of four main stakeholder categories: scholarly publishers, funding agencies, qualitative social scientists and machine learning (ML) researchers, representing a diverse set of perspectives on reproducibility issues and practices. It aims to address the question: **what enablers and barriers do members of these stakeholder communities foresee on the**

road towards a research ecosystem that is conducive to reproducibility?

This study was part of the EU-funded TIER2 project (Ross-Hellauer et al., 2022). TIER2 focused on enhancing reproducibility in research. Its objectives included advancing understanding of reproducibility, developing specialized tools, fostering community engagement, and executing policy interventions in diverse research contexts. It ultimately aimed to improve the reusability and overall quality of research outcomes. A key aspect of the project concerned its investigation of the optimal adaptation of these interventions to the unique epistemic, cultural, and sociotechnical nuances present in various research contexts. The project centered co-creative methods to achieve this goal. This study was preregistered, including a detailed description of its methodology (Tijdink et al., 2023).

This paper empirically contributes to the literature on reproducibility in two ways. It first provides an overview of what the participants of our study consider desirable future states of academia in relation to the reproducibility debate. Second, we describe the enablers and barriers that our participants envision on the road towards these imagined futures. These contributions inform the debate on reproducibility and practically inform researchers, policymakers and other stakeholders on how to transform academic practices and infrastructures to enable community-desired modes of working and foster reproducibility practices.

Methodology and Data

Exploring the Future of Reproducibility Through Future Studies

Our study employs a future studies methodology (Bell, 2009), a systematic approach to exploring possible, probable, and preferable futures. This method aims to envision the future by creating alternative scenarios that inform strategies for shaping desired outcomes at various levels, from individual to global (Boulding & Boulding, 1995; Inayatullah, 2013). By envisioning both desired and undesired futures, futures studies challenge existing frameworks and assumptions, providing a foundation for rethinking the steps necessary to realize or avoid these potential futures. This study was approved by the Graz University of Technology ethical review board on 7th June 2023.

We provide a condensed version of the methodological procedure here.[1]

Methodological Approach: Future Studies and Pre-Workshop Survey

At the core of our method was the creation of different scenarios , designed to explore the future of reproducibility over a ten-year horizon. We used the framework on future studies developed by Sohail Inayatullah (2013) to further refine the methods that would lead to the scenario planning. We engaged four distinct stakeholder categories: ML researchers, qualitative social science researchers, research funders, and scholarly publishers. Funders and publishers were selected as core stakeholders of the TIER2 co-creation community, offering direct interventions to support reproducibility. ML and qualitative social science researchers were selected as opposing ends of the research methodology spectrum, bringing epistemic diversity to our study. Consequently, we organized four dedicated workshops, each tailored to a specific stakeholder group. Participants were recruited based on the following inclusion criteria: People should be working in the specific stakeholder group targeted for the specific workshop and have demonstrable expertise and experience with issues of reproducibility. The latter refers to either studying reproducibility-related issues or contributing to projects or procedures that aim to improve reproducibility, or to tackle related issues.

This future studies process began with a pre-workshop survey, allowing participants to familiarize themselves with the Miro board and the topic of the workshop (the future of reproducibility), the studies' objectives, and the exercises planned (see Supplement 1 for survey questions). This survey collected demographic data and participants' perspectives on the current and future states of reproducibility in their fields or profession, feeding this data into the workshops.

Each workshop was designed to include diverse perspectives from different stakehold-

[1] Greater detail is provided in the preregistered protocol: https://osf.io/wgtku/.

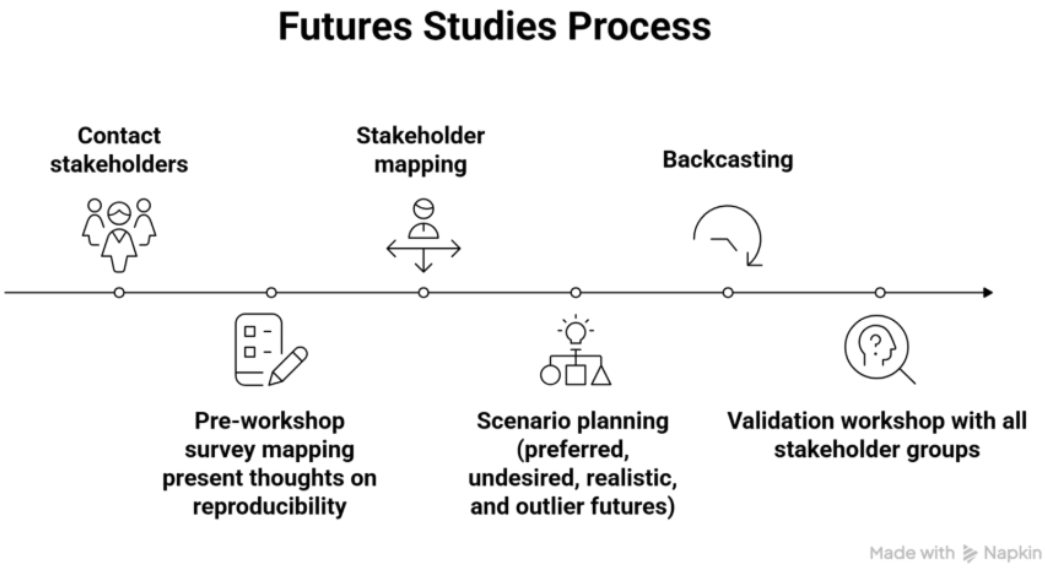

Figure 1 Overview of the methodological framework applied in the futures studies. The figure was produced with the assistance of Napkin AI.

ers. The workshops were conducted online in late June 2023 over 3.5 hours and were divided into three substantive sessions, facilitated by Joeri Tijdink, PhD, MD (male) using Miro, with various members of the authorship team present to support each workshop. Dr. Tijdink is trained as a psychiatrist and meta-researcher and has ample experience in moderating workshops and focus groups. These sessions included a variety of exercises structured around a diverge-converge analysis model, encouraging the individual or pairwise generation of numerous ideas, followed by group reflection, and subsequent selection of the most promising scenarios for further development. Exercises ranged from mapping current reproducibility practices to scenario planning, where participants envisioned desired, undesired, realistic and out-of-the-box scenarios for reproducibility.

The workshops consisted of 3 phases with different aims. The aim of the first phase was for participants to become familiar with thinking about the future by first reflecting on the present, especially through identifying important themes that could impact reproducibility and identifying the main stakeholders that could foster reproducibility practices in the future. The second phase consisted of different scenario planning exercises to discover the desired, undesired, realistic, and out-of-the-box futures of reproducibility. The third phase was centered around backcasting, a specific goal for reproducibility, with the aim of identifying concrete steps different stakeholder groups can take to work towards the preferred futures of reproducibility. This also included identifying main barriers and enablers that will emerge on the road towards these future scenarios.

For more details on how we applied Inayatullah's future studies framework, we refer to the preregistration of this study. There, we also describe the methods used in greater detail and provide the rationale for focusing on the three main pillars of Inayatullah's framework. The 6 pillars are the basis of his general theoretical framework for futures thinking that is linked to methods and tools. These pillars are 1) mapping, 2) anticipation, 3) timing, 4) deepening, 5) creating alternatives, 6) transforming. As proposed by Inayatullah (2013) we adapt this general framework to the conditions and objectives of the TIER2 project, resulting in the approach described below. In our study, we will use 3 pillars. First, we map the current state of reproducibility (1st pillar), then deepen the future by creating alternatives (5th pillar) and start exploring the path to transformation (6th

pillar) by deepening our ideas about facilitators and barriers on the road to transformation.

See also Supplement 2 for our workshop facilitation guide and Supplement 3 for the template Miro board.

Overall, the workshops employed participative and co-creative methods, aimed to gather diverse insights and collaboratively develop comprehensive scenarios, and contributed to a deeper understanding of the (possible) future landscape of reproducibility in research. The workshop series was concluded with a validation event, to which all participants were invited. During this event, we presented initial findings and adjusted based on participant feedback.

Recruitment, Informed Consent and Sample

A pool of potential participants was generated by the research team based on their existing professional networks and the stakeholders already involved in the TIER2 project, within which this study was carried out. Our aim was to generate a sample that was diverse in terms of gender, age, experience, and geography. We recruited participants using a template email initially sent by a member of the team that was affiliated, or networked, with the stakeholder group in question (see Supplement 4). Recruitment was carried out from May through June 2023. Out of 67 people invited, we were able to recruit a sample of 19 participants (see Table 1). Participants were informed about the study and their participation in it through an online form. Participants were presented with an information letter prior to granting informed consent to participate in this study (see Supplement 5). Despite our efforts to recruite a gender-balanced sample, our sample included more men than women (12 vs. 7). It was diverse in terms of age and experience, and in terms of current geographic affiliation, with 9 countries represented overall (8 within Europe, and the United States).

Data and Analysis

All workshop audio and video were recorded in Zoom, transcribed by an automated transcription tool (Amberscript), and checked for accuracy by members of the team. In addition to these transcriptions, our data comprises the responses collected through the pre-workshop survey and information created on the Miro boards during the workshops. All data were compiled in NVivo for coding and analysis. Data were first organized in NVivo using a flexible coding strategy (Deterding & Waters, 2021) to facilitate coding and analysis within and across discrete aspects of the workshops by author NLC (as defined and illustrated in Supplements 2 and 3). We then deductively and collaboratively created a code system based on our organization of the workshops and existing knowledge of their contents (see Supplement 6 for our deductive code system). Then, each member of the team carried out coding of an assigned section of the study using this system. *This codebook was structured around the main themes that were discussed in the workshop.* Members had the freedom to inductively create codes as themes emerged through this process. All data coded by individual team members were merged and four members (NLC, SH, SK and TRH) were then assigned specific themes to focus on for analysing, synthesising and reporting data for the purposes of this paper.

Results

Key Stakeholders and Aims?

In the initial stakeholder mapping workshop exercises, almost all participants reported that researchers are the most influential stakeholders shaping the future of reproducibility. They agreed that what happens in terms of reproducibility is ultimately up to researchers, because it is they who manifest cultures of research. One participant pointed out that researchers are central because they play a variety of roles that can influence the development of reproducibility, including as reviewers and collaborators. Other participants added that professional societies, domain-specific committees and international committees are relevant stakeholders, which are also typically composed of researchers. As one participant from the funders workshop put it: "If researchers don't engage with reproducibility, then we will achieve nothing."

Most participants agreed that funders are the second most influential stakeholder group, placing them centrally in a ring of concentric

	Funders	Qualitative researchers	Publishers	Machine learning	Totals
Number of participants	4	5	5	5	**19**
Countries represented	3	5	3	3	**9**
Men	3	2	4	3	**12**
Women	1	3	1	2	**7**

Table 1 Participant demographics

circles used for this mapping layer (see page five of our Supplement 3, our Miro board template), primarily in the second layer of influence. Exceptions here were ML researchers, who placed them in the center, alongside themselves, and funders, who placed themselves slightly outside of other secondary stakeholder groups. They mostly agreed that publishers have the same level of influence. There was consensus among participants that these two groups are influential because it is their policies, standards, and practices of assessing, rewarding and recognizing research that have the power to influence the development towards research cultures that include reproducibility. We provide more details on this in our findings on enablers of reproducibility. Participants believed that funders have a lot of influence on researchers, because researchers will comply with their demands to get funded. Indeed, we observed during our validation workshop, when presented with the results of this exercise, that some participants asserted that funders should be considered equally influential to researchers because of their power to influence how researchers behave.

Other stakeholder groups, which were largely considered to be secondarily influential, alongside funders, include research institutions and institutional policymakers. Participants also mentioned infrastructure and service providers, libraries, research integrity offices, multi-stakeholder organizations, students and educators. Participants viewed governments as less influential, mostly placing them in the second layer removed from the center, although a few placed them in the first layer and a few in the third. Yet, as a participant from the funder workshop pointed out, governments, when making science policy, set key performance indicators (KPIs) for funders and other stakeholders, and therefore they play an important role in the pipeline of creating incentives for reproducible research.

These findings from our stakeholder mapping exercises are reflected in the discussions that took place across all four workshops. Our results showed that researchers are the stakeholder group that were mentioned most frequently in discussions (116 coded mentions), along with funders (86 coded mentions) and publishers (79 coded mentions). Others, considered less influential in the stakeholder mapping process, received far fewer mentions.

In the scenario planning exercise, participants described their preferred futures of reproducibility as dynamic and identified various characteristics of them, which can be grouped under four main themes: culture, standardization, incentives, and infrastructure. Though we distinguish them here for analytical purposes, participants stressed that they are intertwined and interdependent, while new issues will arise with time and progress. According to our participants, a preferred future has (some of) the following characteristics (see Table 2).

The stakeholders' preferred futures showed significant similarities but also some differences, particularly in the definition of reproducibility. Qualitative researchers emphasized the need for enough variability in qualitative research methods, highlighting that "one-size-fits-all" does not apply to their preferred futures. ML researchers shared similar concerns for computational reproducibility. Publishers and funders, in contrast, focused on the role of standards, guidelines and incentives in fostering reproducible research.

Enablers and Barriers

In brief, the findings we report in this subsection reveal that cultural and social aspects,

Culture	• Research quality more important than quantity • Reproducible practices are centered • Reproducibility part of core training for researchers
Standardisation	• Minimum reproducibility requirements are present among funders, publishers and other stakeholders • Standardized and shareable methods and statistical software with detailed instructions are available to research communities • Guidelines account for epistemic and methodological diversity
Incentives	• Incentives are present to foster reproducible and open research practices • Recognition and visibility of reproducible actions • Alternative research outputs are rewarded • Collaboration is fostered • Reproducible and open research practices are supported by funding
Infrastructure	• Guidelines and policies for using infrastructures are clear • Open-source tools that support reproducible and open research are widely available

Table 2 Four themes of preferred reproducibility futures, with examples

alongside technological and infrastructural ones, are thought to be the most important factors facilitating or hindering both preferred futures and reproducibility more generally. Research culture is conceived of as the space where reproducibility is understood and implemented, as well as the space that must be influenced by other facilitators for reproducibility to become mainstream, normative, and standardized to the extent possible. Meanwhile, technological and infrastructural aspects are understood to be the tools, systems and platforms with, and within which, researchers can implement reproducibility.

To achieve reproducibility in both realms, participants identified training and education (for researchers and all other stakeholders involved in support and assessment), systemic and policy-related factors (to drive and enforce best practices), and financial or economic factors (in terms of funding for reproducible research and for the work and infrastructures that support it) as important factors. In the subsections below, we discuss the barriers and enablers associated with these dimensions in more detail.

Cultural and Social Enablers and Barriers

- **Research Culture: Norms, Values and Practices**

In terms of how it affects reproducibility, our participants framed culture as both the 'place' where reproducibility does or does not happen, and as an enabler of, or barrier to, it. It is both the space where reproducibility is or is not understood and implemented, as well as a set of values, norms and practices that must be influenced by other enablers for reproducible research and workflows to become mainstream, normative and standardized to the extent possible.

Respondents illuminated the enabling role of research culture by pointing out how reproducibility practices sometimes spill over from one area of research to another through publication trends or collaboration. For example, a researcher in our ML workshop said:

> *But it's true that [in] machine learning [...] it was very common from many years ago to just publish the code. So, you have the source code. And for us in recommender systems, until the machine learning community appear[ed] and started publishing a lot, we didn't really pay a lot of attention to that. (ML researcher, scenario planning)*

In our workshop with qualitative researchers, one said of collaboration:

> *Like, even if I'm doing qualitative work, there might be other people doing all sorts of different work and interdisciplinary settings where they might very clearly incentivize me to do different things, or in order to be part of something, some sort of larger project [one has to adopt reproducibility practices]. (Qualitative researcher, scenario planning)*

Conversely, both groups of researchers identified poor levels of awareness about reproducibility and its importance as a barrier to it. In our pre-workshop survey, one ML researcher suggested that those immersed in such issues may exist in a "bubble", while others remain relatively unaware, pointing to the existence of differing and distinct research cultures.

Reflecting a desired shift in research culture, participants in our funders workshop imagined that, in their ideal future of reproducibility, it would be both mainstreamed and normalized. They imagined that reproducibility will be "the new normal", where reproducibility is practiced both "by default" and "by design". A qualitative researcher imagined that researchers are doing reproducible research because "all their peers are choosing to use reproducible methods and approaches", highlighting the potential for research culture to act as an enabler for reproducibility. Elaborating on an imagined future wherein there is "intrinsic motivation" to engage in reproducible research, they added that qualitative researchers are practicing reproducible methods "just because they want to do it," because "it's what's done."

In contrast, a ML researcher suggested during the scenario planning exercise that aspects of research culture that are extrinsic could play an important role. For example, they imagined a future in which each lab has a dedicated staff member to check the work of others by reviewing code and checking for alignment in results reporting. They elaborated this this would shift focus from quantitative metrics of evaluation to quality of the work.

The final point made by this ML researcher signals another theme that we observe in how participants discussed changes in research culture that would foster reproducible research—changes in values. Primarily, participants who spoke about this described a shift in values from quantity of research outputs to quality of research processes and practices. In our workshop with qualitative researchers, participants framed this as a paradigm shift to "slow science". One said:

> *And then this is very crucial. There should also be something of a paradigm shift towards slow science as the way of doing things well, which we believe is going to be the final facilitator to get to the scenario where this sector-wide review is going to find that [reproducibility] is all over the place just because people really, really enjoy it and it brings them good science. (Qualitative researcher, backcasting)*

- **Institutional Culture: Norms, Values and Practices**

Participants observed that for reproducibility to be normalized, it must be embedded in the cultures of institutions and assessment. This differs from how culture was discussed in the previous section, in that the former refers to research cultures made up of communities of peers within a field, discipline or research area, while the latter refers to the culture within a single organization or institute. Note that individuals are always members of both types of cultures at the same time and both types of cultures interact. Maybe owing to the absence of participants directly responsible for institutional cultures of reproducibility, we observed participants mainly voicing barriers rather than enablers in relation to this theme.

Referring to the importance of institutional culture, a funder said, "Because we deem that reproducibility is important, institutions and policymakers also need to embed that in their culture saying, 'You know, this is important to us'." While, in the same workshop, another funder said, "Our future hope is that reproducibility is fully embedded in how research is being evaluated by funding councils, not as an add-on, but really normalized within the standard procedures." Later, the same funder clarified that, in their ideal future of reproducibility, it would be included in a "broadened" set of practices that are "recognized and rewarded… So

that provides space for reproducibility in there. It's maybe not focused on reproducibility, but it allows that to be in."

Participants observed that institutional cultures and practices that foster competition rather than collaboration, and that reward "high impact" publications rather than rigorous reporting of results, are detrimental to reproducibility. In a workshop, a publisher reported anecdotal evidence of groups intentionally withholding methods to prevent replication and linked this specifically to their perception of the current research landscape as one of "competition as opposed to collaboration, sort of as a foundation of why research is done and how research is done." Reflecting the same concern, a funder noted that the dystopian scenario they had formulated (of over-competition and sensationalism in science communication) was in fact quite reflective of the current reality:

> *It's weird that some of this dystopian, negative thing is the extension of how things currently are, [...] forcing a really negative culture that forces over-competition, loss of reputation from reproducing stuff. Researchers are only trained to chase the new, leading them to kind of exaggerate[e] every little discovery in science communications where the public thinks, you know, everything is going to cure cancer and change the world and solve, you know, global warming and stuff. But none of it does because it's not reproducible. It's all just shiny, flashy, new things which causes a loss of trust and loss of faith in science. (Funder, scenario planning)*

Participants in the qualitative researchers workshop also discussed this point but turned responsibility for it back onto funders. Imagining a future in which values have shifted among funders, with a switch from focusing mostly on "sexy topics" that generate new and greater quantities of research outputs, to "funding replication" (backcasting), a qualitative researcher said, "So the novelty value of topics is sort of not the only value for why we are funding things for research, but also to look at these more methodological issues" (backcasting).

On the possibilities for reform, participants were mixed. While participants were all very clear on the need to realign incentives to achieve cultural change towards open and reproducible research, they noted a range of difficulties including the general difficulty of disincentivizing "shortcuts" and "cheating" (Publisher), and resistance from elite researchers prospering within the current system. As a funder observed in the funders workshop, "Why do they push back? It devalues their work." Superficial commitment to DORA (the San Francisco Declaration on Research Assessment)—i.e. not using journal-based bibliometric indicators for assessing individuals—was noted by a publisher as an example of the difficulty of changing engrained habits: "Most people say that the institutions pay lip service to DORA. I would say that almost every scientist I ever encounter judges people on the journals in which they publish."

Participants also expressed concerns regarding possible negative side-effects of overly rigorous reforms expressed through requirements, especially that they could be reduced to merely "ticking the box" (Qualitative researcher, scenario planning). Further still, an ML researcher stated during scenario planning that requirements might be rejected by researchers if they are seen as overly onerous or limiting of creativity and diversity in research: "Many of the scientists I know, they are really creative and playful people. And if you put too much rigor on these researchers, they will not enjoy research and start doing something else."

We note that discussions about cultural norms and values as enablers of or barriers to reproducibility were mainly discussed in the workshops with researchers, rather than funders or publishers. In addition, and potentially consequently, most of the actions mentioned as being required or desired relate to researchers.

- **Standardization of Definitions, Norms and Practices**

Participants viewed lack of consensus on terminology related to reproducibility as a key barrier. Firstly, the terms "reproducibility" and "replicability" themselves are sometimes used synonymously or sometimes given broad or specific meanings. As stated in response to the pre-workshop survey by publishers, "reproducibility" can have a broad or narrow meaning: "It is both a specific thing (computational repro-

ducibility - same data + same code = same results), and an umbrella term that includes consistency of results." In the survey an ML researcher expressed that there is a need to be clearer about which aspects of research are the target of reproducibility, "Are we talking about reproducibility of experiments, of empirical observations, or are we talking about reproducibility of the method? Is that method something computational and all these kinds of things?"

Confusion or lack of clarity about such issues was noted by participants as hampering common understanding. As a researcher put it in the ML workshop during the discussion of survey results, "Every paper about reproducibility, it starts with saying reproducibility is a cornerstone of science and then we don't agree on what it is." One publisher advised, during the discussion of survey results, that such confusion can derail discussions of specifics like the effort and cost associated with ensuring reproducibility. Within qualitative research, the very language of reproducibility was an issue for participants. As we examine later, the relevance of reproducibility to such work is highly contentious. Hence, just using the terms 'reproducibility' and 'replication' can act as a barrier to engaging qualitative researchers in broader discussions of the benefits of greater transparency, which could create possibilities for secondary analysis of data. This was communicated in a discussion during the backcasting exercise with qualitative researchers, wherein one participant stated that conversations about reproducibility are off-putting to qualitative researchers, but that "new questions" can be pursued from shared qualitative data, which they viewed as having value. These findings suggest that standardizing one particular definition of reproducibility would be neither appropriate nor effective as an enabler, however, standardizing other things, like epistemically diverse practices that enable reproducibility, and what is offered and expected by infrastructure and services, would be helpful. In our workshop with qualitative researchers, one spoke in depth about the need for this kind of standardization.

> *In this future scenario, I imagine that there will be standardized methods, workflows that are shareable and standardized. For example, instructions on how to share step-by-step qualitative analysis, because that's also important for reproducibility and qualitative researchers just don't know how to share analysis. And I was talking to this one researcher at some point, and they were like, 'Oh yeah, I'm recording video walkthroughs of, for example, how I code in NVivo my data.' And I'm like, 'Oh, wow, that's amazing, right?' If something like that can be shared. So, this sort of, you know, more innovative maybe, and more standardized ways of sharing these types of outputs so that it's normal and natural and easy for qualitative researchers to share these types of outputs. (Qualitative researcher, scenario planning)*

- **Epistemic Diversity**

Building on the results reported in the previous subsection, our participants agreed that standards should be field-specific and context-sensitive for them to be appropriate and effective. Reflecting this, a publisher said:

> *If you work in this field, this is what you have to do. This is the accepted norm. And I think that, you know, we can do that on a field-specific basis, like people, you know, in certain areas, clinical trials, for example, you understand you would pre-register your hypothesis. You know, if you're doing a certain type of neuroscience experiment, you understand these are your, the fMRI parameters you should be using. So, you have all those standardization. (Publisher, scenario planning)*

Similarly, a qualitative researcher discussed standards specific to qualitative research that would be developed in an ideal future of reproducibility:

> *Metadata standards have been developed. So, if you are providing some sort of contextual appendices or any other type of data that goes with your raw qualitative data that you are archiving and how to do that. And guidelines were also developed on how to reconcile consent and replicability across qualita-*

> *tive approaches. (Qualitative researcher, backcasting)*

Describing how epistemic diversity could be reflected in standards and requirements, a qualitative researcher described a flexibility to data-sharing expectations during scenario planning and said:

One project can share some sort of derived data and another project can share the sort of the analysis workflow or the step-by-step analysis description; that there are these different outputs depending on the type of the study that can be shared. (Qualitative researcher, scenario planning)

Reflecting the same concern, one funder stated in the pre-workshop survey, "I tend to agree with [Sabina] Leonelli's observation that commentators on the importance of reproducibility and its associated terms (replicability, repeatability) tend to overlook these differences and sometimes overemphasize its value as proxy for quality and reliability in science."

Participants observed that, historically, certain quantitative-oriented fields have led in addressing issues of reproducibility while others have thematized the issue less, due to a perceived lack of relevance or importance. As expressed by qualitative researchers, even the term 'reproducibility' can have a marginalizing effect. As one expressed in the pre-workshop survey, "I don't think reproducibility (even broadly construed) is a goal for the majority of qualitative research", while another responded in the survey:

> *The concept of reproducibility as defined, even in the broadest sense, does not really hold as a meaningful and/or desirable concept for many working in the traditions of qualitative research—although note there are some more realist perspectives for whom the concept has some grounding.*

Another qualitative researcher shared in the survey, reflecting the marginalizing effects of the current reproducibility discourse and movement:

> *For the majority of members of the qualitative psychology community the drive for reproducibility presents a threat/risk. Wholesale moves towards centering reproducibility at the heart of what counts as 'good' science risks eroding/ devaluing research where reproducibility is not a measure of methodological rigor or research quality.*

Therefore, for reproducibility reforms to be relevant and valued across diverse research epistemologies, it is important to know what the term means and how it is used, and to recognize that other terms might be more applicable and resonant for some.

- **Leaders in Change**

Our participants envision that their preferred futures of reproducibility will be enabled by community-driven standards that reflect and respect epistemic diversity. Some conceived of this in a more grass-roots fashion, while others suggested that a top-down approach, led by professional associations, groups of international experts and multi-stakeholder organizations should be pursued. Reflecting the discussions about research culture and institutional culture as enablers or barriers regarding reproducibility, the discussions in this section highlight different views among our participants about which of these two types of cultures is best positioned to lead the way towards a desirable future of reproducibility.

Describing a community-driven approach but also suggesting that a higher-level organization could foster this, a ML researcher said:

> *So, I think the community [...] should really meet and start agreeing on things because, even if you make a paper, that has a limited impact, right? So, I think the community or maybe [professional organization] or maybe someone at a higher level should not directly set a standard but try to gather the community [...]. A higher-level institution like ACM [Association for Computing Machinery] should put together researchers and start finding agreements and I would say if we make this, then start a little by little [...]. Then in calls for papers or reviewer criteria. (ML researcher, backcasting)*

A funder also spoke to the importance of research communities as leaders in spreading reproducibility norms and practices (often open science practices) when describing the func-

tion of reproducibility networks. They said, "We recently provided funding for a national reproducibility network in [country]. We think that's a very promising instrument to raise awareness and bring together people working on this topic. We've been very much inspired by network[s] that already exist in, for instance, Germany and the UK, I think. So yeah, I think having those kinds of networks and ensuring that researchers can work on that together [is important]" (Scenario Planning).

> *Reflecting the idea that professional societies should have a leading role in defining standards, a publisher noted during the stakeholder mapping exercise that they had "moved my professional societies closer to the center" because of the role they play in certain fields of setting standards that research must abide.*

While professional societies tend to be composed of researchers, they have authority and power that is markedly different from an independent research community. Therefore we understand this view to be different from those stated previously.

In a similar vein, some participants suggested that international experts and multi-stakeholder organizations should lead in standard-setting. One publisher suggested, during stakeholder mapping, that, when international committees of academics/experts define and set standards, these become the norm within the relevant communities, and these will often be taken up and enforced by publishers, hence flagging them as a potential enabler for reproducibility.

Systemic and Policy Issues

A second factor discussed in all workshops concerns the implementation of adequate policy to foster reproducibility. This can relate to a host of different practices, ranging from data sharing, transparent reporting, appointing data stewards, providing training, etc., and it involves various actors, though mainly the traditionally powerful ones (funders, publishers, institutions, or even national governments) as they are best positioned to enforce certain practices from actors that depend upon them. Participants gave various examples in which they "only made progress" when certain policies were mandated, or projected expectations about what would happen if certain powerful actors would start enforcing or demanding reproducibility practices. For example, a funder said during the stakeholder mapping exercise, when pointing out that governments are important stakeholders, "If the government says institutions, funders, we want you to focus on reproducibility and this is what we value, then people will do it."

Although generally agreed upon for their efficacy and necessity, mandates spurred debate due to the need for universal standards among all stakeholders to prevent irrelevant or infeasible tasks. Yet, permitting exceptions was understood to undermine a policy's effectiveness. This tension became evident in the ML researchers' workshop backcasting exercise, highlighting policy's dual role as both an enabler and a barrier in achieving reproducible futures. One participant first proposed that a paper would be poorly evaluated without the accompaniment of documentation that supports reproducibility, but then backtracked because "not all research needs to be reproducible." In turn, this prompted their exercise partner to state, "But if you keep it for researchers to choose, most of the researchers will not choose to do it because it's extra work."

Another argument in favor of implementing and mandating policies across the board is that it can make implementation easier because it removes 'first mover risks'. These could be detrimental especially for researchers (being afraid of not being able to get their work funded or published) and publishers (being afraid of losing submissions). Participants stressed that this speaks for stakeholders to act in collectives, rather than as individual entities. They suggested, for example, that associations of publishers, rather than individual publishers, are better placed to act

A recurring theme in relation to effective implementation of policies is the establishment of adequate incentive structures. Participants across the four workshops saw the need to integrate reproducibility practices into the full system and lifecycle of assessment. For example, a publisher stated in the pre-workshop survey, "If we want to improve the situation of reproducibility, it needs to be integrated in how research groups are awarded grants, fol-

lowed up during a project, or measured in their institutions."

While during the backcasting exercise, a funder commented:

I think the thing that needs fixing is the whole incentive structure. [...] in the future, I think research assessment should be based on the quality of the research rather than the impact. And it should be judged on whether the methodology is good. Has there been preregistration? Things that actually help improve trust in science. (Funder, Backcasting)

Related to this, institutions were particularly mentioned and perceived to be important facilitators of reproducibility practices. They do not have a primary role in conducting reproducible research, but they are considered to be important supporters of reproducibility practices. They can do so either through properly rewarding and incentivizing such practices, or by providing the institutional conditions to achieve reproducibility, e.g. through training, installing data stewards or providing adequate infrastructure. Participants acknowledged that incentive structures are particularly important, which is partly a duty of institutional actors. However, one funder shared during the stakeholder mapping exercise, that in a particular national context, they didn't perceive institutions to play an important role in this process.

Participants note that there is a need for some level of standardization across contexts to remove barriers like 'first mover risk'. However, as we mentioned before, they also recognized a risk of moving too much into a 'one-size-fits-all model', which comes with its own limitations and is considered an important barrier for widespread uptake of diverse reproducibility practices. Institutionalization of reproducibility norms on a discipline level was by some considered to provide a suitable balance.

Technological and Infrastructural Enablers and Barriers

Participants discussed the significance of suitable infrastructures and technological advancements as enablers of reproducibility practices, and the absence thereof as barriers. They agreed upon and emphasized the essential nature of having proper infrastructure for reproducibility in place. Some mentioned the need for reproducibility tools and public infrastructure, while not specifying concrete features. Others mentioned the importance of data sharing and curation, highlighting the need for additional resources to help store, share and reuse data. The level of detail provided in discussions seemed to relate to the degree of participants' engagement in developing or using specific infrastructures or tools. Those more actively involved in such processes were more likely to provide detailed accounts of how technological and infrastructural elements could act as either enablers or barriers of reproducibility practices.

Infrastructures or tools to share and reuse data were discussed most. Participants in all workshops mentioned that these are required to promote reproducibility. However, participants in multiple workshops also mentioned that infrastructures for data sharing and reusing on their own are insufficient. For example, in the publisher workshop, one respondent commented on connecting research items (e.g., data, papers, code) by linking platforms, tools, and repositories. A particular challenge named in this context was the missing possibility of linking between scientific platforms, partly attributed to a lack of agreements on meta-data standards and guidelines.

The need for instruments (tools, protocols, routines) supporting the whole workflow towards reproducible science was also discussed. Participants brought attention to challenges such as the lack of agreements on metadata standards and guidelines, and the need for instruments supporting the entire workflow towards reproducible science. For example, a publisher said during a workshop:

So if there's going to be lots of different research objects out there online in different places, like you might have papers in one place, data in one repository code in another, that all link to other grants, like all of the linking of all the different objects, needs to be clearer and better to just stay discoverable and therefore reproducible further down the line. (Publisher, stakeholder mapping)

Additionally, the discussion emphasized the need for publicly funded infrastructures to support reproducibility on an institutional level. Connecting to another type of barrier discussed previously, a lack of standardization was identified as a prominent issue affecting dissemination, communication, and evaluation methods and tools. In the pre-workshop sur-

vey, one qualitative researcher identified the problem this way: "Different software used for data management and analyses (software can affect data processing and analysis and consequently the results; proprietary software makes reproducibility difficult due to inaccessibility)."

ML participants also pointed out that meaningful code sharing requires more than just having an infrastructure in place, emphasizing the need for appropriate reward mechanisms for checking the validity of code and having dedicated staff for this task.

Finally, the discussion touched upon the improper use of AI technologies, its potential impact on the scientific landscape, and its trustworthiness. Particularly in qualitative research, participants name the uptake of new methods as essential to counteract the high workload of scientists. However, one concern discussed among publishers, funders, and qualitative researchers is the error-prone or improper application of AI technologies (e.g., to fake results). Imagining a dystopian future for reproducibility, one qualitative researcher identified a barrier to reproducibility by expressing concern about "sloppy" use of AI technology to auto code qualitative data. For example, "you read through four [transcripts] and then you tell ChatGPT to do the rest of the coding for you." Noting that proprietary QDAS (Qualitative Data Analysis Software) already have auto-coding features, they continued, "It's terrible for qualitative research, but it's going to happen and it's going to be a mess to review and to disentangle."

Financial and Economic Factors

Our participants overwhelmingly framed financial and economic factors as important enablers of reproducible research and as key barriers to it. They stated that funders can drive reproducible research practices by rewarding them with funding and by requiring them in grant agreements, or hinder them by not doing so. Furthermore, participants stated that funders can foster the spread of values, practices, and awareness by funding certain tracks of research; that they can specifically fund reproduction and replication studies; and that they and other stakeholders can fund the work and initiatives that support reproducible research (e.g., data stewards, trainers, etc.). When this supporting work is not funded, reproducibility is difficult to achieve in the context of a rushed, publish-or-perish research culture.

To this latter point, one qualitative researcher referred to "the cost of reproducibility" in the pre-workshop survey, and explained that extra time and resources are needed to enable transparency of methods, data, code, etc. In the words of one funder, "It is still often costly, sloppy and takes time". An ML researcher noted that this means that not all researchers are able to enact reproducibility:

> *Reproducibility for now is not for everyone. Not everyone is doing reproducibility and not everyone can do it because of resources and costs. Not everyone has the money and funding to do it. (ML researcher, backcasting)*

Framing this as a systemic problem, one funder stated in the pre-workshop survey that reproducibility practices are "currently not really rewarded or recognized within funding streams and processes, including pre-award and post-award processes." While having financial implications, this point relates to wider incentive and assessment structures that participants recognized as potential enablers of or barriers to reproducibility practices.

However, participants believe that funders can solve this problem by specifically funding the labor costs associated with this work. One funder stated during scenario planning that, in an ideal future of reproducibility, funded researchers would be:

Expected to hire data stewards, and these data stewards ha[ve] to work together to collect data according to their principles. And if they wouldn't agree to work together in that scheme that we set up, they just wouldn't get the money for the research. So, I think if we would want to as funders, there's ways to do this.

In a similar vein, a qualitative researcher commented during scenario planning:

One would wish there was like grant funding on top of, not just as part of what is already offered. You could actually have people as part of your project who are actually taking care of all the reproducibility issues, right?

They continued, "And that means that there is work hours that have to be put into this and

that people are being hired to do this right with you or for you."

In terms of the research content that is funded, participants spoke about the concept of funders creating traction or "critical mass" around reproducibility and related issues, like the funding stream that supported the TIER2 project and others like it. One ML researcher commented during backcasting, "So, the fact that we have this meeting and there's research projects funded by EU, there is some traction in this." And a funder commented similarly during scenario planning:

And then of course, all of these projects like TIER2 and others will continue to be moving, going forward. Right? There'll be more of them. And so, the awareness and critical mass around reproducibility will hopefully continue and increase until something bigger happens. (Funder, scenario planning)

Finally, some participants, including funders, suggested that funders should specifically fund reproduction and replication studies to effectively "kickstart" these practices. A qualitative researcher supported this approach and suggested it as an alternative to funders focusing on "sexy topics". They stated that if funders put out calls for such studies, then researchers would respond.

Funders, meanwhile, recognized both pros and cons to this suggested approach:

> *I agree with most of what you're saying, but something that struck me was this funding for reproducibility research that you brought up? Yeah, I agree. I agree with you. There should be more. But if we're thinking about ideal scenarios, like in ten years, and I put a comment here, maybe, you know, it should just be embedded everywhere, right? So, it shouldn't have specific funding for it. (Funder, Scenario Planning)*

However, one funder noted that, despite an interest in supporting reproducibility, "It's harder for us to ask for [reproducibility] because of the way that we fund things, because it's less stable and more project based." Thereby, the funder suggests that supporting reproducibility requires longer-term funding and too short programs could act as barriers to reproducibility.

Another type of economic barrier identified by our participants is the proprietary nature of some data. Within our ML workshop, economic issues regarding data as intellectual property arose in two contexts by the same participant. In general, collaboration with industry partners for whom "data is a digital gold, and they don't want to have it everywhere but actually with them and not sharing" was seen to limit the data sharing that ensures reproducibility, during the backcasting exercise. The same respondent also cited a more specific case, namely the issue of research done on proprietary datasets where changes in licensing conditions can mean the data on which previous research rests is no longer available. ML researchers noted that industry-funded research also presents a barrier, in that sharing of data, software and other materials is often not allowed by industry partners for reasons of competition. Researchers working in such collaborations would not be able to create reproducible research.

Training and Education

In addition to the previously discussed dimensions, training and educational activities are considered crucial for promoting and scaling reproducibility practices. Participants in all four workshops stated that formal training, mentorship, and role modelling are essential to "get policy down to the individual researcher" and to "incorporate it in everyday research" (ML researcher, backcasting), making "those things normal" (Publisher, Scenario Planning).

> *Participants suggest that training can address multiple aspects of research and involve various actors. "The primary focus could be on early career researchers or students, teaching them technical skills, research methods and statistics, as well as open science skills" (Publisher, Survey). Several participants experienced such training (either as trainer or trainee) and described how these are essential elements of a desired future, "[…] the reproducibility practices are integrated also in the university curriculum, so that from early stages on, students who will become researchers at some point are already familiar with these practices" (Funder, backcasting).*

Some participants took it a step further by

suggesting that PhD students should be required to conduct a replication study to obtain their degree. Additionally, senior researchers, principal investigators, and others in various roles, such as evaluators at journals or funding agencies, should also receive training.

Especially among qualitative researchers, this was even extended to non-academic staff at universities, including support staff facilitating tasks related to open science e.g., data management and analysis. One participant suggested that efforts to teach about reproducibility could go beyond academia to raise public awareness about the importance of reproducibility, aligning with traditional science literacy models. Participants emphasized the need for training to go beyond early career researchers and students, and to not only focus on research practices or data-related skills. They believe that training should lead to a cultural shift, where researchers understand that the additional effort required to make their studies reproducible will be rewarded.

On the other hand, some participants expressed concerns that improper implementation of role modelling or mentorship could hinder reproducibility. One funder highlighted the risk of early career researchers being socialized into a culture that normalizes cutting corners to attain funding (backcasting).

In addition, participants discussed the need for resources to support these training and mentorship efforts. One funder suggested that funding agencies could provide dedicated schemes or resources for these activities, possibly embedded within existing funding schemes for training reproducible practices (scenario planning).

Discussion and Conclusion

This study set out to identify the enablers and barriers that members of four key stakeholder communities (scholarly publishers, funders, qualitative researchers, and machine learning researchers) foresee on the way toward a desired future state of the research ecosystem that fosters reproducibility. We found that enablers and barriers can be categorized into five main clusters. The factors most prominently mentioned as potentially supporting or hindering a desired future are located within research culture, including norms, values and shared definitions. Additionally, they are located in the infrastructure required to engage in reproducibility practices, including repositories, support staff, and digital infrastructure required for sharing research materials. Three other clusters of factors put forth by participants relate to policy efforts required to incentivize reproducibility practices; training and education to empower researchers and support staff to engage in reproducibility practices; and the financial resources required to facilitate the transition towards a desired future and to specifically fund replication studies. The main enablers and barriers identified by study participants are summarized in Table 3.

The future research system that participants imagine:

- Has a particular research culture that prioritizes quality over quantity and centers reproducibility in research practice and in training;
- Has standardized reproducibility requirements that account for methodological and epistemic diversity, and standardized and shareable methods, tools and workflows;
- Incentivizes reproducible, open, and collaborative practices by providing them with recognition, funding, and visibility (these include alternative research outputs); and
- Has infrastructure that is designed for ease of use with clear guidance, policies and training; hosts FAIR and open tools and workflows; and is backed by sufficient resources to develop and maintain such infrastructure.

Our results generally align with previous studies that have assessed the implementation of open science practices (e.g., Drude et al., 2022) and reproducibility (e.g., Lalu et al., 2023), indicating a need for a culture change and training for all actors involved to achieve the desired goal.

As mentioned before, several of the themes or topics discussed by the participants transcend the boundaries of the analytical categories used for our discussions. For example, research assessment and incentive structures relate to multiple categories. Moreover, some of the topics were highlighted both as enablers and potential barriers. We note a few tensions between and across the enablers and barriers identified by participants. First, there remains ambiguity about the level of standard-

Clusters of factors	Enablers	Barriers
Social and cultural	- Research cultures in which reproducibility practices are the norm	- Insufficient awareness of reproducibility practices and their importance
	- Inclusion of reproducibility practices in institutional reward and recognition	- Lack of consensus on terminology related to reproducibility
	- Recognition of the diverse contexts related to reproducibility across research fields and traditions	- Enforcement of too rigid reproducibility standards across all research contexts
	- Community-driven leaders in change, preferably including actors from diverse stakeholder groups	
Systemic and policy	- Sufficiently flexible policy available mandating open/reproducible practices where relevant	- Lack of policy or overly rigid policy that leads to irrelevant or infeasible tasks
	- Policy incentivizing collective action to avoid 'first mover risks'	- Lack of proper reward systems for reproducibility practices
Technological and infrastructural	- Suitable infrastructures and tools available for reproducibility practices	- Lack of standardization in tools and infrastructure
	- Availability of open-source software and tools	- Improper use of AI technologies
Financial and economic	- Sufficient resources available to implement reproducibility practices and the labor costs involved	- Too few resources available for additional work or infrastructure required for reproducibility
	- Funding available for replication studies	- Proprietary nature of some data/intellectual property right issues hindering open and reproducible practices
Training and educational	- Training to make reproducibility practices part of daily research work	- Improper implementation of role modelling or mentorship
	- Performing replication study as part of PhD training	
	- Training for those doing evaluation	

Table 3 Overview of main enablers and barriers towards a desirable future research ecosystem in relation to reproducibility practices as identified by study participants

ization or flexibility that should be maintained in the pursuit of reproducibility. Discussions amongst participants, addressing both the infrastructures and cultural aspects, exemplified the pros and cons of adopting a common set of standards shared across researchers and other stakeholders in diverse disciplines and contexts. These were contrasted with a more flexible approach, catered to the specific needs of diverse communities, potentially involving distinct approaches in different settings. Thus, echoing the work of Leonelli (2018, 2022), our findings suggest great value in context-sensitive solutions and expectations that respect the diversity of research practices and epistemologies.

Linked to this are questions of ownership and collaboration: to what extent, and on what scale, should stakeholders join forces to address reproducibility standards? Participants recognized the need for cross-cutting approaches involving different stakeholders from different communities. However, they also recognized the risk of ignoring the specificities of individual research contexts if efforts are coordinated too centrally. The merits of bottom-up versus top-down innovations and standards were debated. Therefore, the feasibility and desirability of developing standards, tools and guidelines in collaboration with different actors remained ambiguous. All in all, we echo our participants' view that researchers, with the many hats that they wear, should be in leading positions to develop these. However, guidance from institutional actors would be useful in setting basic standards and expectations, in linking and syncing with infrastructures and services, in centering epistemic diversity, and in fostering community-driven initiatives. Researchers in their roles as members of reproducibility networks and referees in grant or manuscript review, or as members of institutional review boards, are well-positioned to contribute to this agenda. This could create a dialectic of providing structure while empowering creativity and specificity at the local level.

This accommodation of structure and creativity relates to the discussions about the desirability of mandating reproducibility efforts, and if so, against which standards. Some participants forcefully argued that the only way to achieve higher levels of reproducibility was by mandating efforts to foster it, for example, through an explicit requirement in grant funding, journal publication or tenure processes. Others, however, were much more skeptical, arguing for the need to remain flexible. This skeptical view included maintaining the option to opt-out of reproducibility efforts and standards, in cases where these are not deemed relevant or feasible. This finding supports recent trends in the literature, which frame reproducibility (and replicability) as neither universally applicable nor feasible across diverse epistemic contexts (Cole et al., 2024; Leonelli, 2018; Penders et al., 2019; Ulpts & Schneider, 2023). Consequently, we recommend the development of guidelines for reproducibility (and/or transparency) practices tailored to specific domain, methodological and epistemic contexts.

In relation to discussions about infrastructure, we make two additional observations that connect the workshop discussions to wider concerns in the community. First, we note that in most cases, participants refer to infrastructures that they either actively use themselves or have the possibility to develop. This was, for instance, the case of a so-called knowledge graph discussion by the publishers or a data stewards discussion in the funder workshop. Hence, participants put a certain level of responsibility on their shoulders by discussing the need for infrastructures that they can or should develop. Second, we note that infrastructure was mentioned not only as a direct enabler of or barrier to reproducibility practices, but also in connection with the changing temporalities of science. In this sense, the proper use of infrastructures and tools was thought to allow more effective use of time, which could ultimately benefit reproducibility practices and research quality more generally. Reference was made to recent developments in AI, which could free up time for qualitative researchers.

The results of this study speak to broader, ongoing conversations in the literature about the importance of research culture, in either fostering positive outcomes regarding reproducibility, or inhibiting them (Uttley et al., 2025). Our participants flagged, as have many others, that the fast-paced, time-poor nature of research that is tied to funding cycles and the "publish-or-perish" mentality, often stands in the way of rigorous and reproducible research. For example, a study focused on the

prevalence of questionable research practices (QRPs) found that publication pressure appears to increase the frequency at which researchers engage in QRPs (Gopalakrishna et al., 2022). Reflecting research culture, evidence suggests that mentoring plays a role in shaping how junior researchers operate. The same study from Gopalakrishna et al. (2022) found that "survival mentoring" (i.e. the practice of teaching mentees how to survive in academia) is associated with increased rates of QRPs, while "responsible mentoring" exerts a (weaker) influence in reducing the rate of QRPs.

Some of our participants recognized the tension between a hurried research culture and research rigor, integrity and reproducibility. They therefore called for a shift to "slow science" – a solution that has been posited by others in recent years (Frith, 2020; Leite & Diele-Viegas, 2021; Stengers, 2018). According to Frith (2020), to do slow science, one would need to produce less but better research; orient to longer timescales and bigger horizons; shift to valuing quality over quantity; value collaboration; and teach and mentor in ways that reflect these values. This is echoed by some of our participants. Importantly, this view raises the issue of research assessment and its role in shaping research culture. As our participants pointed out, what is valued in assessment procedures—be they within research institutions, funders, or publishers—wields a heavy influence on the values, norms and practices of research cultures. If reproducible research practices are not valued in assessment, and the time that is required to implement them is not accounted for in metrics used for assessment purposes, then reproducible research will not be realized. Related to this are concerns about the time and financial cost of creating and sharing reproducible research, which, though extensive for all, are not borne evenly. They are borne more so by early career researchers (Schimanski & Alperin, 2018) and by women, in the case of data management (Larivière et al., 2021). This has implications not just for inequity in the undervaluing of this work in assessment procedures, but also in terms of which researchers are better positioned to implement them, depending on the level of resources available at their home institutions (Ross-Hellauer, Reichmann, et al., 2022).

Inequity in reproducibility labor and differences in implementability call for approaches to reproducibility that focus on the full research ecosystem and lifecycle, including assessment procedures as a prime lever for initiating transformation. Transforming assessment procedures arguably creates both the opportunity and motivation for scholars to engage in reproducibility. The Behaviour Change Wheel framework (Michie et al., 2011) suggests these are two of the three main influences on behavior—together with capability, which can be fostered through relevant training and infrastructure.

In terms of the latter, we agree with our participants' view that initiatives are needed that focus on equipping researchers, and support staff, with the necessary skills to engage in reproducibility practices, starting from early-career stages. These must be community-defined, -driven and -delivered to be effective. Echoing participants, we recommend that training begin within the research education system.

In conclusion, some enablers and barriers listed in Table 3 directly translate to specific recommendations for diverse stakeholders. For example, the enabler of 'funding being available for replication studies' translates into the recommendation for research funders to design programs that are open to replication studies. We will not list all such direct translations into recommendations here for brevity's sake. We do make three main recommendations that capture the main enablers and barriers in slightly less obvious ways:

- Research institutes and funders: Develop context-sensitive solutions and expectations for reproducibility that reflect the diversity of research practices and epistemologies
- Researchers, research funders and publishers: Foster community-driven initiatives for developing infrastructures and reproducibility and/or transparency guidelines tailored to specific domain, methodological, and epistemic contexts
- Research institutes and funders: Recognize the additional resources required for reproducibility practices (e.g. in terms of labor, skills, infrastructure, and tools), making sure these are available and fairly distributed among the actors involved.

In sharing these conclusions and making these recommendations, we acknowledge the limitations of this study. With 19 participants across four stakeholder categories, the sample size per stakeholder group is relatively small. And, while diverse in some respects, our pool of participants was limited in terms of geographic, cultural, and gender representation. Participants were predominantly from Europe and the United States, with a higher representation of men than women. In addition, participants were selected for their expertise and experience in relation to reproducibility, thereby likely engaging those that feel reproducibility is important. This limited diversity likely affected the range of perspectives included in the study, particularly regarding how reproducibility challenges are perceived in different global research contexts. We further acknowledge that the stakeholder workshops were conducted online and within a limited time frame, which may have influenced the depth and quality of the discussions.

Despite the limitations, our findings confirm other findings already established in the literature and add to it by synthesizing perspectives from a diverse stakeholder community, leading to actionable recommendations to transition research cultures and infrastructures towards a future state of research that fosters reproducibility, where and when appropriate.

Acknowledgements

We are grateful to our study participants for dedicating the time and effort to engage in our study and providing invaluable insights. We also want to express our gratitude to Eva Kormann for their help in preparing this manuscript.

Funding

This work was supported by the project TIER2, funded by the European Union's Horizon Europe research and innovation program under grant agreement No. 101094817.

Supplementary Material

Supplements referred to in this manuscript are available at: https://osf.io/n28sg/

References

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Bazzoli, A. (2022). Open science and epistemic pluralism: A tale of many perils and some opportunities. *Industrial and Organizational Psychology*, *15*(4), 525–528. https://doi.org/10.1017/iop.2022.67

Bell, W. (2009). *Foundations of futures studies, volume 1: Human science for a new era*. Transaction Publishers.

Bennett, E. A. (2021). Open science From a qualitative, feminist perspective: Epistemological dogmas and a call for critical examination. *Psychology of Women Quarterly*, *45*(4), 448–456. https://doi.org/10.1177/03616843211036460

Bergeat, D., Lombard, N., Gasmi, A., Le Floch, B., & Naudet, F. (2022). Data sharing and reanalyses among randomized clinical trials published in surgical journals before and after adoption of a data availability and reproducibility policy. *JAMA Network Open*, *5*(6), e2215209. https://doi.org/10.1001/jamanetworkopen.2022.15209

Betz, C. L. (2011). Adoption of CONSORT statements for randomized control trials published in the Journal of Pediatric Nursing. *Journal of Pediatric Nursing*, *26*(3), 177–178. https://doi.org/10.1016/j.pedn.2011.02.001

Boulding, E., & Boulding, K. E. (1995). *The Future: Images and processes*. SAGE Publications.

Branney, P. E., Brooks, J., Kilby, L., Newman, K., Norris, E., Pownall, M., Talbot, C. V., Treharne, G. J., & Whitaker, C. M. (2023). Three steps to open science for qualitative research in psychology. *Social and Personality Psychology Compass*, *17*(4), e12728. https://doi.org/10.1111/spc3.12728

Cobo, E., Cortes, J., Ribera, J. M., Cardellach, F., Selva-O'Callaghan, A., Kostov, B., Garcia, L., Cirugeda, L., Altman, D. G., Gonzalez, J. A., Sanchez, J. A., Miras, F., Urrutia, A., Fonollosa, V., Rey-Joly, C., & Vilardell, M. (2011). Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: Masked randomised trial. *BMJ*, *343*, d6783–d6783. https://doi.org/10.1136/bmj.d6783

Cole, N. L., Ulpts, S., Bochynska, A., Kormann, E., Good, M., Leitner, B., & Ross-Hellauer, T. (2024). Reproducibility and replicability of

qualitative research: An integrative review of concepts, barriers and enablers. OSF. https://doi.org/10.31222/osf.io/n5zkw

Crüwell, S., Apthorp, D., Baker, B. J., Colling, L., Elson, M., Geiger, S. J., Lobentanzer, S., Monéger, J., Patterson, A., Schwarzkopf, D. S., Zaneva, M., & Brown, N. J. L. (2023). What's in a badge? A computational reproducibility investigation of the open data badge policy in one issue of *Psychological Science*. *Psychological Science*, *34*(4), 512–522. https://doi.org/10.1177/09567976221140828

Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative Health Research*, *12*(2), 279–289. https://doi.org/10.1177/104973202129119793

de Rijcke, S., & Penders, B. (2018). Resist calls for replicability in the humanities. *Nature*, *560*(7716), 29–29. https://doi.org/10.1038/d41586-018-05845-z

Deterding, N. M., & Waters, M. C. (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological Methods & Research*, *50*(2), 708–739. https://doi.org/10.1177/0049124118799377

Drude, N., Martinez-Gamboa, L., Haven, T., Holman, C., Holst, M., Kniffert, S., McCann, S., Rackoll, T., Schulz, R., & Weschke, S. (2022). Finding the best fit for improving reproducibility: Reflections from the QUEST Center for Responsible Research. *BMC Research Notes*, *15*(1), 270. https:/ / doi.org/ 10.1186/ s13104-022-06108-x

Drummond, C. (2019). Is the drive for reproducible science having a detrimental effect on what is published? *Learned Publishing*, *32*(1), 63–69. https://doi.org/10.1002/leap.1224

Eich, E. (2014). Business not as usual. *Psychological Science*, *25*(1), 3–6. https://doi.org/10.1177/0956797613512465

Evans, T. R., Branney, P., Clements, A., & Hatton, E. (2023). Improving evidence-based practice through preregistration of applied research: Barriers and recommendations. *Accountability in Research*, *30*(2), 88–108. https://doi.org/10.1080/08989621.2021.1969233

Fanelli, D., Costas, R., & Ioannidis, J. P. A. (2017). Meta-assessment of bias in science. *In Proceedings of the National Academy of Sciences*, *114*(14), 3714–3719. https://doi.org/10.1073/pnas.1618569114

Fidler, F., & Wilcox, J. (2021). Reproducibility of Scientific Results. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/

Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences*, *24*(1), 1–2. https://doi.org/10.1016/j.tics.2019.10.007

Fuller, T., Peters, J., Pearson, M., & Anderson, R. (2014). Impact of the transparent reporting of evaluations with nonrandomized designs reporting guideline: Ten years on. *American Journal of Public Health*, *104*(11), e110–e117. https://doi.org/10.2105/AJPH.2014.302195

Giofre D., Boedker I., Cumming G., Rivella C., & Tressoldi P. (2022). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices: Five years later. *Behavior Research Methods*, *17*, 3845-3854. https://doi.org/10.3758/s13428-022-01993-3

Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PloS One*, *17*(2), e0263023. https:/ / doi.org/ 10.1371/journal.pone.0263023

Guttinger, S. (2020). The limits of replicability. *European Journal for Philosophy of Science*, *10*(2), Article 10. https://doi.org/10.1007/s13194-019-0269-1

Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., deMayo, B. E., Long, B., Yoon, E. J., & Frank, M. C. (2021). Analytic reproducibility in articles receiving open data badges at the journal *Psychological Science*: An observational study. *Royal Society Open Science*, *8*(1), 201494. https://doi.org/10.1098/rsos.201494

Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. P. A. (2022). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). *Perspectives on Psychological Science*, *17*(1), 239–251. https://doi.org/10.1177/1745691620979806

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

Haven, T. L., Errington, T. M., Gleditsch, K. S., Van Grootel, L., Jacobs, A. M., Kern, F. G., Piñeiro,

R., Rosenblatt, F., & Mokkink, L. B. (2020). Preregistering qualitative research: A Delphi study. *International Journal of Qualitative Methods*, *19*. https://doi.org/10.1177/1609406920976417

Holbrook, J. B., Penders, B., & De Rijcke, S. (2019). The humanities do not need a replication drive. *CWTS*. https:/ / www.cwts.nl:443/blog?article=n-r2v2a4&title=the-humanities-do-not-need-a-replication-drive

Hopewell, S., Boutron, I., Altman, D. G., Barbour, G., Moher, D., Montori, V., Schriger, D., Cook, J., Gerry, S., Omar, O., Dutton, P., Roberts, C., Frangou, E., Clifton, L., Chiocchia, V., Rombach, I., Wartolowska, K., & Ravaud, P. (2016). Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: Results of a randomised controlled trial. *BMC Medicine*, *14*(1), Article 199. https://doi.org/10.1186/s12916-016-0736-x

Huma, B., & Joyce, J. B. (2022). 'One size doesn't fit all': Lessons from interaction analysis on tailoring Open Science practices to qualitative research. *British Journal of Social Psychology 62*(4), 1590-1604. https://doi.org/10.1111/bjso.12568

Inayatullah, S. (2013). Futures Studies. Theories and Methods. In N. Al-Fodhan (Ed.), *There's a Future. Visions for a Better World* (Vol. 30, pp. 36–66). T. F. Editores.

Karhulahti, V., Branney, P., Siutila, M., & Syed, M. (2022). A primer for choosing, designing and evaluating registered reports for qualitative methods [Preprint]. *MetaArXiv.* https://-doi.org/10.31222/osf.io/2azkf

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr., R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142–152. https:/ / doi.org/ 10.1027/ 1864-9335/a000178

Lalu, M. M., Presseau, J., Foster, M. K., Hunniford, V. T., Cobey, K. D., Brehaut, J. C., Ilkow, C., Montroy, J., Cardenas, A., Sharif, A., Jeffers, M. S., & Fergusson, D. A. (2023). Identifying barriers and enablers to rigorous conduct and reporting of preclinical laboratory studies. *PLOS Biology*, *21*(1), e3001932. https://doi.org/10.1371/journal.pbio.3001932

Larivière, V., Pontille, D., & Sugimoto, C. R. (2021). Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies*, *2*(1), 111–128. https://doi.org/10.1162/qss_a_00097

Leite, L., & Diele-Viegas, L. M. (2021). Juggling slow and fast science. *Nature Human Behaviour*, *5*(4), 409–409. https://doi.org/10.1038/s41562-021-01080-1

Leonelli, S. (2018). Rethinking reproducibility as a criterion for research quality. In Fiorito, L., Scheall, S., & Suprinyak, C. E. (Eds.), *Including a symposium on Mary Morgan: Curiosity, imagination, and surprise* (pp. 129–146). Emerald Publishing Limited. https://doi.org/10.1108/S0743-41542018000036B009

Leonelli, S. (2022). Open Science and epistemic diversity: Friends or foes? *Philosophy of Science*, *89*(5), 991–1001. https://doi.org/10.1017/psa.2022.45

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, *6*(1), 42. https://doi.org/10.1186/1748-5908-6-42

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 229–237. https://doi.org/10.1177/2515245920918872

Open Data Barometer. (2017). *Global Report Fourth Edition*. https:/ / opendatabarometer.org/ doc/ 4thEdition/ ODB-4thEdition-GlobalReport.pdf

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic

reviews. *BMJ*, *372,* Article 71. https://doi.org/10.1136/bmj.n71

Peels, R., & Bouter, L. (2018). The possibility and desirability of replication in the humanities. *Palgrave Communications*, *4*(1), 1–4. https://doi.org/10.1057/s41599-018-0149-x

Penders, B., Holbrook, J. B., & de Rijcke, S. (2019). Rinse and repeat: Understanding the value of replication across different ways of knowing. *Publications*, *7*(3), Article 3. https://doi.org/10.3390/publications7030052

Percie Du Sert, N., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., Browne, W. J., Clark, A., Cuthill, I. C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S. T., Howells, D. W., Hurst, V., Karp, N. A., Lazic, S. E., Lidster, K., MacCallum, C. J., Macleod, M., … & Würbel, H. (2020). Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2.0. *PLOS Biology*, *18*(7), e3000411. https://doi.org/10.1371/journal.pbio.3000411

Plesser, H. E. (2018). Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics*, *11*. https://doi.org/10.3389/fninf.2017.00076

Pownall M, Azevedo F, Konig L M, Slack H R, Evans T R, Flack Z, Grinschgl S, Elsherif M M, Gilligan-Lee K A, de Oliveira C M. F, Gjoneska B, Kalandadze T, Button K, Ashcroft-Jones S, Terry J, Albayrak-Aydemir N, Dechterenko F, Alzahawi S, Baker B J, … Sadhwani S. (2023). Teaching open and reproducible scholarship: A critical review of the evidence base for current pedagogical methods and their outcomes. *Royal Society Open Science*, *10*(5), 221255. https://doi.org/10.1098/rsos.221255

Pratt, M. G., Kaplan, S., & Whittington, R. (2020). Editorial essay: The tumult over transparency: Decoupling transparency from replication in establishing trustworthy qualitative research. *Administrative Science Quarterly*, *65*(1), 1–19. https://doi.org/10.1177/0001839219887663

Reischer, H. N., & Cowan, H. R. (2020). Quantity over quality? Reproducible psychological science from a mixed methods perspective. *Collabra: Psychology*, *6*(1), 26. https://doi.org/10.1525/collabra.284

Ross-Hellauer, T., Bouter, L. M., & Horbach, S. P. J. M. (2023). Open peer review urgently requires evidence: A call to action. *PLOS Biology*, *21*(10), e3002255. https://doi.org/10.1371/journal.pbio.3002255

Ross-Hellauer, T., & Horbach, S. P. J. M. (2024). Additional experiments required: A scoping review of recent evidence on key aspects of Open Peer Review. *Research Evaluation*, *33*, rvae004. https://doi.org/10.1093/reseval/rvae004

Ross-Hellauer, T., Klebel, T., Bannach-Brown, A., Horbach, S. P. J. M., Jabeen, H., Manola, N., Metodiev, T., Papageorgiou, H., Reczko, M., Sansone, S.-A., Schneider, J., Tijdink, J., & Vergoulis, T. (2022). TIER2: Enhancing trust, integrity and efficiency in research through next-level reproducibility. *Research Ideas and Outcomes*, *8*, e98457. https://doi.org/10.3897/rio.8.e98457

Ross-Hellauer, T., Reichmann, S., Cole, N. L., Fessl, A., Klebel, T., & Pontika, N. (2022). Dynamics of cumulative advantage and threats to equity in open science: A scoping review. *Royal Society Open Science*, *9*(1), 211032. https://doi.org/10.1098/rsos.211032

Rowhani-Farid, A., Aldcroft, A., & Barnett, A. G. (2020). Did awarding badges increase data sharing in *BMJ Open*? A randomized controlled trial. *Royal Society Open Science*, *7*(3), 191818. https://doi.org/10.1098/rsos.191818

Rowhani-Farid, A., & Barnett, A. G. (2018). Badges for sharing data and code at Biostatistics: An observational study. *F1000Research*, *7*, Article 90. https://doi.org/10.12688/f1000research.13477.2

Sarewitz, D. (2016). Saving science. *The New Atlantis*. https://www.thenewatlantis.com/publications/saving-science

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00813

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

Schimanski, L. A., & Alperin, J. P. (2018). The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future. *F1000Research, 7, Article 1605*. https://doi.org/10.12688/f1000research.16493.1

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., Esterling, K. M., & Nosek, B. A. (2021). Initial evi-

dence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, *5*(8), 990–997. https://doi.org/10.1038/s41562-021-01142-4

Stengers, I. (2018). *Another science is possible: A manifesto for slow science* (S. Muecke, Trans.). Polity.

Tamminen, K. A., & Poucher, Z. A. (2018). Open science in sport and exercise psychology: Review of current approaches and considerations for qualitative inquiry. *Psychology of Sport and Exercise*, *36*, 17–28. https://doi.org/10.1016/j.psychsport.2017.12.010

Tijdink, J. K., Leitner, B., Horbach, S., Cole, N. L., Kopeinik, S., & Ross-Hellauer, T. (2023). *Future studies—the future(s) of reproducibility in research* [Preregistration]. https://doi.org/10.17605/OSF.IO/WGTKU

Ulpts, S., & Schneider, J. W. (2023). *Knowledge production modes: The relevance and feasibility of reproducibility* [Preprint]. OSF. https://doi.org/10.31222/osf.io/ujnd9

Uttley, L., Falzon, L., Byrne, J. A., Tricco, A. C., Munafò, M. R., Moher, D., Stoeger, T., Matandika, L., Labbé, C., & Naudet, F. (2025). Research culture influences in health and biomedical research: Rapid scoping review and content analysis. *Journal of Clinical Epidemiology*, *178*, 111616. https://doi.org/10.1016/j.jclinepi.2024.111616

Van Der Braak, K., Ghannad, M., Orelio, C., Heus, P., Damen, J. A. A., Spijker, R., Robinson, K., Lund, H., & Hooft, L. (2022). The score after 10 years of registration of systematic review protocols. *Systematic Reviews*, *11*(1), Article 191. https://doi.org/10.1186/s13643-022-02053-9

Veroniki, A. A., Tsokani, S., Zevgiti, S., Pagkalidou, I., Kontouli, K. M., Ambarcioglu, P., Pandis, N., Lunny, C., Nikolakopoulou, A., Papakonstantinou, T., Chaimani, A., Straus, S. E., Hutton, B., Tricco, A. C., Mavridis, D., & Salanti, G. (2021). Do reporting guidelines have an impact? Empirical assessment of changes in reporting before and after the PRISMA extension statement for network meta-analysis. *Systematic Reviews*, *10*(1), Article 246. https://doi.org/10.1186/s13643-021-01780-9

Zuiderwijk, A., Janssen, M., & Parnia, A. (2013). The complementarity of open data infrastructures: An analysis of functionalities. *Proceedings of the 14th Annual International Conference on Digital Government Research*, 166–171. https://doi.org/10.1145/2479724.2479749