Enhancing Trust, Integrity, and Efficiency in Research through Next-Level Reproducibility

Deliverable 3.2 – Validated key impact pathways for reproducibility, including recommendations

31/01/2026

Lead Beneficiary: Know Center Research GmbH

Author/s: Tony Ross-Hellauer, Eleni Adamidi, Stefania Amodeo, Alexandra Bannach-Brown, Nicki Lisa Cole, Thomas Klebel, Simone Kopeinik, Eva Kormann, Barbara Leitner, Fakhri Momeni, Haris Papageorgiou, Jesper Schneider, Joeri Tijdink, Thanasis Vergoulis

Reviewers: Eleni Adamidi (ARC), Susanna Sansone (UOXF)

Funded by
the European Union

Prepared under contract from the European Commission
Grant agreement No. 101094817

EU Horizon Europe Research and Innovation action

Project acronym:	**TIER2**
Project full title:	**Enhancing Trust, Integrity, and Efficiency in Research through Next-Level Reproducibility Impact Pathways**
Start of the project:	January 2023
Duration:	36 months
Project coordinator:	Dr. Tony Ross-Hellauer
Deliverable title:	Validated key impact pathways for reproducibility, including recommendations
Deliverable n°:	D3.2
Version n°:	1.0
Nature of the deliverable:	Report
Dissemination level:	Internal
WP responsible:	WP3
Lead beneficiary:	Know Center Research GmbH

TIER2 Project, Grant agreement No. 101094817

Due date of deliverable:	31/12/2025
Actual submission date:	31/01/2026

Version	Status	Date	Author(s)
0.1	Outline of deliverable draft	11/09/2025	Nicki Lisa Cole
1.0	First inputs	31/10/2025	Nicki Lisa Cole, Jesper Schneider, Joeri Tijdink, Eleni Adamidi, Thanasis Vergoulis, Eva Kormann, Thomas Klebel, Barbara Leitner, Fakhri Momeni, Stefania Amodeo
1.1	Second round	30/11/2025	Nicki Lisa Cole, Jesper Schneider, Joeri Tijdink, Eleni Adamidi, Thanasis Vergoulis, Eva Kormann, Thomas Klebel, Barbara Leitner, Fakhri Momeni, Stefania Amodeo, Tony Ross-Hellauer
1.2	Full draft	12/01/2026	Tony Ross-Hellauer, Thomas Klebel, Nicki Lisa Cole, Joeri Tijdink
1.3	Review	16/01/2026	Eleni Adamidi, Susanna Sansone
2.0	Final version	31/01/2026	Tony Ross-Hellauer, Eleni Adamidi, Stefania Amodeo, Nicki Lisa Cole, Thomas Klebel, Simone Kopeinik, Eva Kormann, Barbara Leitner, Fakhri

Momeni, Haris Papageorgiou, Jesper Schneider, Joeri Tijdink, Thanasis Vergoulis

Table of contents

Executive Summary

Over the last three years, TIER2 has aimed to address growing concerns about reproducibility in scholarly research by systematically investigating reproducibility in context. We focused on three broad research areas (social, life and computer sciences) as well as two cross-disciplinary stakeholder groups (research publishers and funders). The project has sought to build a conceptual and evidence-based understanding of reproducibility, identify gaps in existing interventions and practices, and develop new tools and frameworks through co-creation techniques such as scenario planning and user-centred design. This work is intended to strengthen the evidence base on reproducibility and to enhance awareness, skills, and community engagement across diverse research communities.

The final phase, represented by this report, aims to summarise and synthesise key findings from TIER2 to reflect upon the role reproducibility plays across different aspects of the research system, and the gains for research quality and integrity that different interventions can bring.

We structure the report around the four Key Outputs from our project:

1. **Enhanced understanding**: Delivered through both a new Conceptual framework for evaluating the relevance and feasibility of reproducibility as a research practice and criterion of research quality in diverse settings, and Evidence synthesis to create a knowledge base on the efficacy of reproducibility interventions. (Sections 2, 3, 4)
2. **Innovative tools and practices**: Delivered through development, implementation, and evaluation of eight highly co-creative pilot activities to create new reproducibility-related tools and practices focused on the social, life, and computer sciences, as well as research publishers and funders. Pilots covered topics from planning tools and reproducible computational workflows to editorial checks, dashboard-based monitoring, and funder-level promotion plans. Our outcomes emphasise the importance of early-stage planning, intuitive infrastructures, cross-stakeholder action and the essential role of research communities. (Section 5)
3. **Increased capacity**: Delivered through the new Reproducibility Hub of resources on the Embassy of Good Science platform, an expansive training course hosted by the OpenPlato platform, creation of three new Reproducibility Networks in Widening Participation countries, and creation of a new platform for meta-research collaboration (MERRI). These activities strengthened skills, awareness, and community infrastructures across Europe (Section 6)
4. **Policy roadmap**: Our co-created final recommendations outline priorities for improving research reproducibility, addressed to researchers, publishers, funders and institutions in Europe and beyond. The recommendations emphasise the need for context-sensitive norms, policymaking that accounts for epistemic diversity across research fields, as well as highlighting the need for stronger incentives, open practices, and dedicated support to enhance the transparency and trustworthiness of AI research. (Sections 7, 8)

List of Abbreviations

API – Application Programming Interface (standard meaning; not expanded in document)
DAS – Data Availability Statement
DEIA – Diversity, Equity, Inclusion, and Accessibility
DMP – Data Management Plan
DO – Digital Object
EOSC – European Open Science Cloud
EU – European Union
FAIR – Findable, Accessible, Interoperable, Reusable
GUPRI – Globally Unique, Persistent and Resolvable Identifier
KPM – Knowledge Production Mode
MERRI – Meta-Research and Reproducibility Infrastructure Collaboration
NLP-OSS – Natural Language Processing Open Source Software
NOADs – National Open Access Desks
OSF – Open Science Framework
PRISMA-ScR – PRISMA Extension for Scoping Reviews
REA – European Research Executive Agency
RMP – Reproducibility Management Plan
RNs – Reproducibility Networks

1. Introduction and methodology

Over the last three years, TIER2 has aimed to address growing concerns about reproducibility in scholarly research by systematically investigating reproducibility in context. We focused on three broad research areas (social, life and computer sciences) as well as two cross-disciplinary stakeholder groups (research publishers and funders). The project has sought to build a conceptual and evidence-based understanding of reproducibility, identify gaps in existing interventions and practices, and develop new tools and frameworks through co-creation techniques such as scenario planning and user-centred design. This work is intended to strengthen the evidence base on reproducibility and to enhance awareness, skills, and community engagement across diverse research communities.

In addition, TIER2 aimed to pilot and implement interventions to improve reproducibility, ensure that developed tools were compatible with broader infrastructures like the European Open Science Cloud (EOSC), and systematically evaluate the effectiveness of these interventions. Finally, the project aimed to synthesise findings into a roadmap of recommendations and policy guidance, with the ultimate goal of increasing the reuse and overall quality of research results and thereby boosting trust, integrity and efficiency in research within the European Research Area and beyond.

The project, as described in our Description of Action (Ross-Hellauer et al., 2022) was underpinned by four guiding principles:

- **Reproducibility is an opportunity, not a crisis:** That rather than framing reproducibility as a crisis, we prefer to view it as an opportunity to rethink how research is done to enable constructive analysis of what works, where, and why across research.
- **Epistemic diversity must be centred:** That reproducibility varies by context. Differences in factors such as control of environments, use of statistics, research aims, interpretation, and technical, social and cultural conditions shape what reproducibility means and when it applies. These variations must be better understood to assess benefits and costs across research and innovation.
- **Evidence must be systematised for informed policy:** That because impacts differ by context, analyses should map how interventions work in practice, including trade-offs and unintended effects. Gains will not be universal, and poorly designed policies may reinforce existing inequalities. Policies should therefore support communities at different stages of development.
- **Action must be targeted and holistic:** That building reproducibility requires aligned efforts: robust infrastructures, practical skills, connected communities, supportive incentives and proportionate policies. Many initiatives already exist; linking and strengthening them is essential.

The project proceeded according to six methodological steps, shown below in Fig 1, whereby early work on conceptualisation and evidence synthesis led into the development, implementation and assessment of eight novel pilot interventions, tools and practices addressing various stakeholders (researchers, funders and publishers).

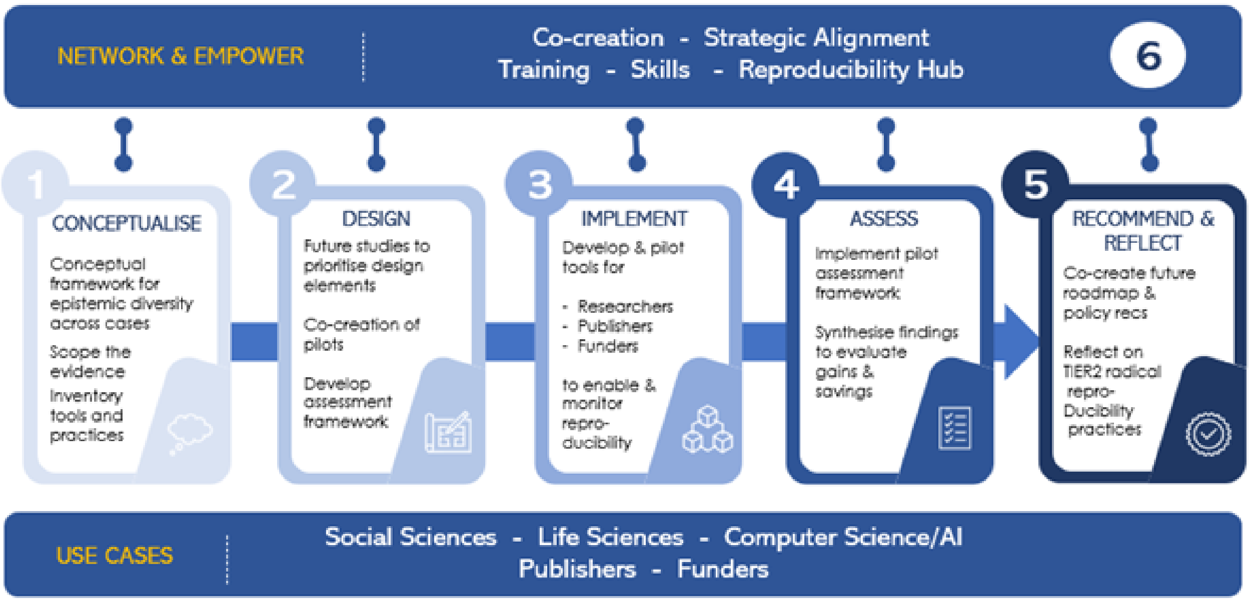

Figure 1. TIER2 methodological steps

The final phase, represented by this report, aims to summarise and synthesise key findings from TIER2 to reflect upon the role reproducibility plays across different aspects of the research system, and the gains for research quality and integrity that different interventions can bring.

We structure the report around the four Key Outputs from our project:

- **Enhanced understanding**: Delivered through both a new Conceptual framework for evaluating the relevance and feasibility of reproducibility as a research practice and criterion of research quality in diverse settings, and Evidence synthesis publications which synthesise the knowledge base on the efficacy of reproducibility interventions. (Sections 2, 3, 4)
- **Innovative tools and practices**: Delivered through development, implementation, and evaluation of eight highly co-creative pilot activities to create new reproducibility-related

tools and practices focused on the social, life, and computer sciences, as well as research publishers and funders. (Section 5)

- **Increased capacity**: Delivered through the new Reproducibility Hub of resources on the Embassy of Good Science platform, an expansive training course hosted by the OpenPlato platform, creation of three new Reproducibility Networks in Widening Participation countries, and creation of a new platform for meta-research collaboration (MERRI). (Section 6)
- **Policy roadmap**: Our co-created final recommendations outline priorities for improving research reproducibility, addressed to researchers, publishers, funders and institutions in Europe and beyond. They are accompanied by a range of policy briefs that focus on promoting reproducibility-sensitive policymaking that accounts for epistemic diversity across research fields, as well as highlighting the need for stronger incentives, open practices, and dedicated support to enhance the transparency and trustworthiness of AI research. (Sections 7, 8)

2. Enhanced understanding (1): A conceptual framework for assessing the relevance and feasibility of reproducibility

Author: Jesper Wiborg Schneider (Aarhus University)

This chapter presents the synthesised conceptual framework developed throughout the course of the project.

2.1. Background

Over the past decade, reproducibility has emerged as a central concern across the academic landscape. Claims of a 'reproducibility crisis' in fields such as psychology and biomedicine have triggered widespread calls for systemic interventions and normative reforms (Baker, 2016). Such calls have led to increasingly vocal demands for near universal reproducibility standards, often codified through journal policies, funding mandates, and institutional best practices (Bissell, 2013; Drummond, 2019). While well-intentioned, many of these policy responses rest on a narrow conceptualisation of science that assumes a single set of standards and criteria can or should apply across all research domains. However, as critics such as Leonelli (2018), Guttinger (2020), and Penders et al. (2019, 2020) argue, reproducibility can take various forms and is not a universally appropriate criterion. Treating reproducibility as universal risks epistemic injustice, marginalising valid forms of inquiry simply because they do not conform to norms imported from experimental or quantitative traditions (Fricker, 2007; Penders et al., 2019). By foregrounding epistemic diversity, the TIER2 project advocates for situated assessments of the appropriateness and applicability of reproducibility as a criterion and practice that considers how specific research practices align with their underlying epistemological commitments and practical constraints.

Policy initiatives often risk misaligning expectations with the actual diversity of research practices and the preferences within different research communities with varying epistemic cultures (Knorr Cetina, 1999). Efforts like the Hong Kong Principles (Moher et al., 2020) or the EU Scoping Review on Reproducibility (European Commission et al., 2020) implicitly promote a model of research rooted in experimental and quantitative traditions. They offer guidelines designed to incentivise practices such as preregistration, data sharing, or (statistical) replication. However, such guidance presumes that reproducibility is always a relevant and feasible indicator of research quality, a presumption that does not hold across the full spectrum of academic disciplines and epistemic cultures. This leads to a structural tension: policies aimed at ensuring research integrity can end up privileging certain kinds of science while disadvantaging others, particularly in qualitative, exploratory, or interpretivist domains. Therefore, without explicit recognition of epistemic diversity, these policies may lead to epistemic injustice or even exclusion by penalizing researchers simply because their legitimate methods do not align with dominant reproducibility norms (Fricker, 2007; Penders et al., 2019).

2.2. Defining reproducibility in terms of '*enabling*' and '*redoing*'

As a first task, we undertook a review of 400+ meanings of reproducibility taken from the literature across all main research fields (Ulpts & Schneider, 2025a). The review work produced a number of insights: 1) the concept takes on various forms and serves different purposes depending on modes of knowledge production. 2) Terminology and definitions are more confused than hitherto indicated by the literature, both within and across disciplines. 3) The body of literature addressing the role and place of 'reproducibility', as well as its connection to epistemic diversity, underscores that a uniform approach to promote and increase ‘reproducibility’ would advantage certain kind of research and researchers for whom such an approach is appropriate, and disadvantage other kinds of research and researcher for whom it is inappropriate.

However, a key distinction that emerges from this conceptual review of the literature is between two crucial underlying meanings about the verb forms *reproduce* and *replicate,* namely that they refer both to 1) practices involved in redoing and 2) the epistemic functions intended to be achieved by these practices. Practices are determined by what parts of a study are to be redone, what should be kept the same or similar and what should be varied. Such parts are manifold and most likely impossible to enumerate in a fixed typology. Our review indicates that dominating discourses often narrowly link epistemic functions of reproducibility or replication to issues such as reliability or validity. However, our review also suggests that such discourses fail to acknowledge the vast number of functions identifiable in the literature. Complexity or diversity is set aside, probably because such taxonomies have normative aims rather than descriptive ones.

We hence propose a practical solution by 1) substituting terms for the acts commonly associated with ‘replication’ and ‘reproducibility’ (and related terms) with *redoing,* and 2) ‘replicable’ and ‘reproducible’ with enabling. We further acknowledge the intricate complexities in describing these practices and their diverse epistemic functions. We suggest that practices and functions are mapped for individual cases while specifying the intended sameness and variation of the key parts. This is a situated definition of a specific kind of *redoing* or *enabling.* The distinction between *redoing* and *enabling* clarifies what is to be done (the practice) and for what purpose (the function), facilitating researchers or stakeholders in navigating this conceptually confused territory.

Read more:

- Ulpts, S., & Schneider, J. W. (2025a). A conceptual review of uses and meanings of reproducibility and replication (Version 3). *MetaArXiv*. https://doi.org/10.31222/osf.io/entu4_v3

2.3. TIER2 KPM framework: The relevance and feasibility of reproducibility

To address the complex relationship between reproducibility and epistemic diversity the TIER2 project has developed a pluralistic and context-sensitive conceptual framework that acknowledges the effect of epistemic diversity on the relevance and feasibility of reproducibility as a criterion and practice for good research. The Knowledge Production Mode (KPM) framework allows for the assessment of the relevance and feasibility of reproducibility across the diverse research landscape based on social, epistemic and contextual conditions of the research (Schneider et al., 2024; Ulpts & Schneider, 2025b).

We hence aim to extend and build upon previous attempts at typologizing reproducibility in the context of epistemic diversity (e.g., Leonelli, 2018; Penders et al., 2019) to offer an account that avoids using flawed organising constructs like disciplines or methods, and accounts for underlying epistemologies that fundamentally shape what counts as legitimate knowledge within an epistemic community.

Therefore, in TIER2, we propose knowledge production modes (KPMs) as an organising construct. A KPM represents the epistemic, social, and contextual conditions under which research is produced and evaluated. A discipline or field can contain several different KPMs, and one KPM can embrace several preferred methodical techniques. The framework shifts away from one-size-fits-all ideals toward an evaluative model, instead focusing on two dimensions of the appropriateness of reproducibility. These dimensions link the epistemic "why" with the practical "how", offering a more refined means of assessing the appropriateness of reproducibility in diverse contexts based on KPMs. Accordingly, first, we must establish whether reproducibility is epistemically **relevant** for the research. If so, we should then examine how **feasible** it is.

Is reproducibility relevant? **Relevance** is assessed by examining underlying epistemic traditions, which determine research goals, epistemic functions, and systems of justification that govern quality criteria and evaluative standards. In other words, they define ways of knowing within a community organised around a KPM (Ulpts & Schneider, 2025b). Therefore, the question is: *Does reproducibility serve a meaningful epistemic function within a KPM, such as establishing reliability or demonstrating generalizability?* Different epistemic traditions (e.g., positivism, constructivism, pragmatism) generate different kinds of knowledge claims and apply different evaluative standards (Carter & Little, 2007; Lincoln & Guba, 1985; Tuval-Mashiach, 2021). For positivist approaches, reproducibility signals rigour; whereas for interpretivist ones, credibility, reflexivity, or plausibility may be more relevant epistemic criteria than reproducibility (Guba and Lincoln, 1985; Tuval-Mashiach, 2021).

Is reproducibility feasible? **Feasibility** concerns practical constraints. It depends on features of the subject matter, such as its complexity and plasticity (Guttinger, 2020), as well as the degree of tacit expertise needed for the task. It also depends on the available resources available and the necessary investments in infrastructure, time, and labour, and by whether materials are proprietary. Finally, the feasibility is affected by theoretical and methodological uncertainty associated with KPM, the subject matter and study goals (Whitley, 2000). Therefore, the question is: can *reproducibility realistically be achieved under these conditions, given the practical, methodological, technical, and epistemic constraints of the research setting.*

Read more:

- Ulpts, S., & Schneider, J. W. (2025). Knowledge Production Modes: The Relevance and Feasibility of Reproducibility (Version 2). *MetaArXiv*. https://doi.org/10.31222/osf.io/ujnd9_v2

2.3. Implications for stakeholders

Unfortunately, good intentions and interventions may backfire if they do not consider the relevance and feasibility of reproducibility for different research contexts.

Stakeholders may, for example, wish to incentivise certain Open Science practices, due to their assumed relationship to reproducibility, but should respect variations in knowledge production and justification. Journals and editors, for instance, often implement data-sharing mandates without considering variation in feasibility caused by differing research properties and conditions. Incentives or mandates rooted in a narrow understanding of science can lead to misalignment between community internal epistemic preferences and practical conditions on the one hand and incentives that reward conformity over quality, on the other. In the worst case this can lead to the exclusion of valid knowledge due to inappropriate standards. But it will most definitely lead to frustration and forms of disengagement from some research communities, particularly in the social sciences and humanities.

To the contrary, policies should be designed to recognise and accommodate epistemic diversity. This means being explicit about what is meant by 'reproducibility' in a given context, acknowledging different epistemic functions, and asking whether the function is relevant, and the practice is feasible for the specific research involved. Practices can either be specific acts of redoing parts or the whole of a study with varying degrees of similarity compared to the previous research or practices of enabling that allow others to either redo the research or just comprehend and trace what was done and why.

As highlighted in the first TIER2 Policy Brief (Schneider et al., 2024), these insights are not just conceptual; they translate directly into policy design questions:

- Which reproducibility policies can be broadly applied, and which should be targeted to specific areas only?
- At what level of granularity should policies operate (e.g., funders, disciplines, journals)?
- How can new reproducibility policies complement existing open science and integrity initiatives?

In TIER 2, the KPM framework was not applied as a formal analytical tool or procedural guide. Rather, it functioned as a heuristic and conceptual lens that informed the formulation of research questions, the interpretation of findings, and our reflections on the epistemic and practical conditions of reproducibility.

2.4. The way forward

In an era of increasing demands for research transparency and rigour, reproducibility remains a vital – but highly contextual – concept and practice. Policymakers play a critical role in ensuring that efforts to promote reproducibility enhance, rather than undermine, research quality and inclusivity. We therefore emphasise the importance of approaching reproducibility with care, recognising that universal mandates risk overlooking epistemic diversity and the varied ways in which knowledge is produced across the diverse research landscape.

Good practice means tailoring reproducibility approaches to specific research communities, ideally in collaboration with those communities themselves, and being explicit about the purposes and practical meanings of reproducibility within guidelines and mandates. Where reproducibility is not relevant or not feasible, attention to alternative epistemic functions such as transparency, reflexivity, and traceability, is essential. By embracing a pluralistic, pragmatic, and principled approach, stakeholders can foster a healthier research ecosystem that respects the diverse ways knowledge is made, justified, and shared.

3. Enhanced understanding (2): The evidence base for efficacy of interventions to improve reproducibility

This chapter puts our conceptual framework in conversation with the literature base that describes impact pathways for reproducibility. In scoping the evidence base, in line with our emphasis on epistemic diversity, TIER2 sought width and breadth of investigation. Hence, we focused our efforts firstly on scoping the evidence base for evidence of the efficacy of interventions across disciplines (Section 3.1), and then on investigating two distinct methodological areas with rich implications for reproducibility, Machine Learning (Sec. 3.2) and qualitative research (Sec. 3.3).

3.1. The evidence base for reproducibility interventions across disciplines

Author: Eva Kormann (Know Center)

In a joint study with colleagues from our sister project OSIRIS, we used the PRISMA-ScR methodology (Tricco et al., 2018) to systematically scope the literature for evidence of the efficacy of interventions to improve reproducibility and replicability across disciplines (Dudda et al., 2025). Searching ten databases and screening over 36,000 records led to just 105 tested interventions from 86 empirical articles containing relevant evidence, most of them from health and behavioural research and published in recent years. Figure 2 presents an overview of the main areas of investigation of each study.

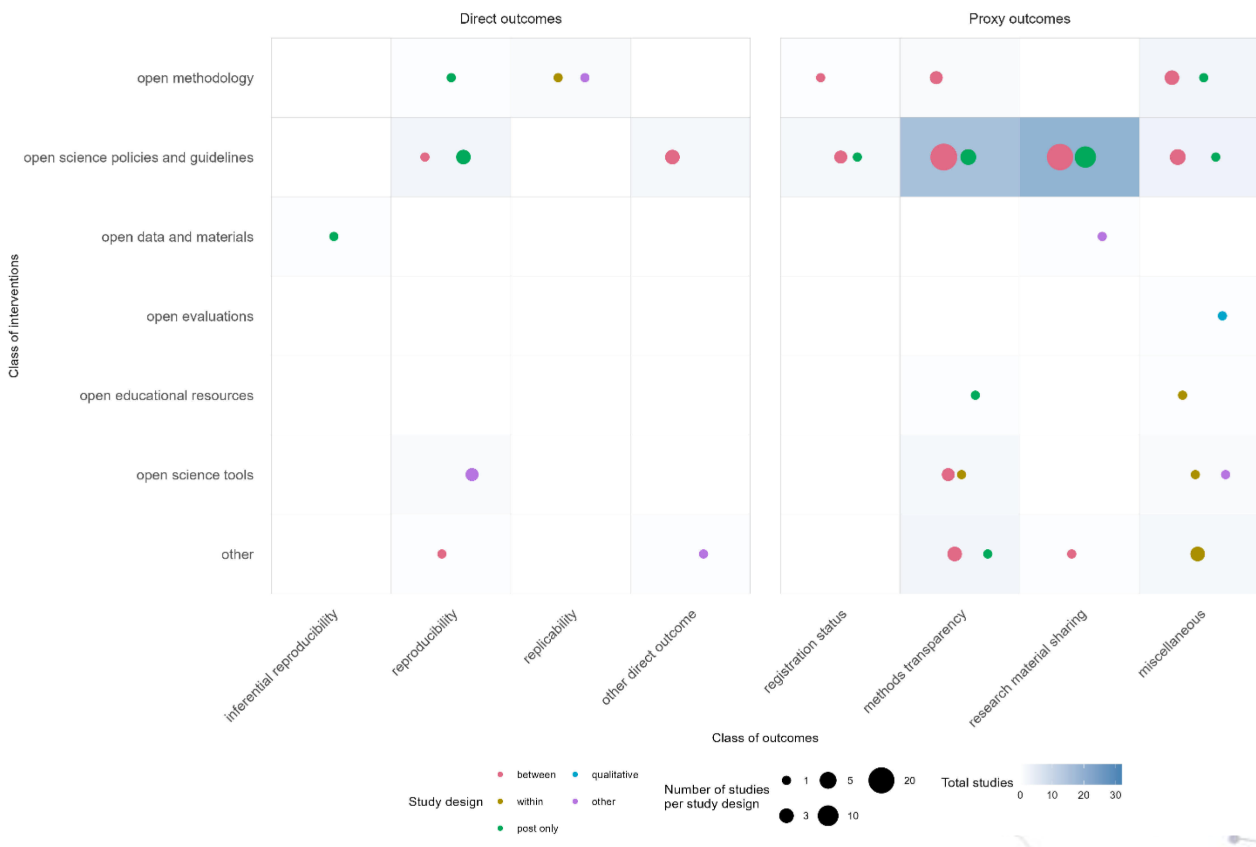

Figure 2. Evidence gap map of interventions and outcome domains investigated, with information on study designs. The left pane shows direct reproducibility outcomes, while the right pane shows proxy outcomes. The size of the bubbles refers to the number of studies that address this particular intervention/outcome combination; the colours refer to different study designs.

As can be seen in the evidence gap map, findings demonstrate that very few interventions have so far been tested which target reproducibility or replicability as a direct outcome. This evidence was mainly be identified related to computational reproducibility, i.e., rerunning computational analyses to obtain the same results. A small share of studies were concerned with replicability, i.e., repeating (parts of) studies obtaining the same results. The main findings are briefly summarized in Table 1.

Table 1. *Summary of main findings from scoping review on direct outcomes*

Direct outcome	Intervention	Findings	Exemplary references
Computational reproducibility	Software tools	Help to make code re-executable	(Chan & Schoch, 2023)
	Publisher data sharing mandates	Show positive effects, but reproducibility still depends on other factors such as completeness of materials or code availability	(Bergeat et al., 2022)
	Registered reports	Increase accessibility of materials, but have limited effect on reproducibility	(Obels et al., 2020)
	Badges	Increase accessibility of materials, but have limited effect on reproducibility	(Hardwicke et al., 2021)
Replicability	Methods standardisation	Improves replicability across study sites	(Arroyo-Araujo M et al., 2019)

More than investigating reproducibility or replicability directly, we found studies concerned with outcomes that could be considered their proxies, mainly transparency of methods and data sharing (i.e., *enabling* rather than *redoing* according to our definitions in the previous section). In particular, the effect of reporting guidelines has been extensively studied, in addition to various publisher policies (especially those aiming to increase sharing of research objects). The main findings are compiled in Table 2. Further evidence, e.g., on badges, generally suggests that incentives alone are not very effective in increasing reproducibility practices. The effectiveness of mandates also depends on various factors, such as how strongly they are enforced or which standards are already present in a certain field. Studies on other practices and outcomes are scarce.

Table 2: *Summary of main findings from scoping review on proxy outcomes*

Proxy outcome	Intervention	Findings	Exemplary references
Methods transparency	Reporting guideline publication	No improvements observed after mere publication of guidelines	(Palmer et al., 2021; Veroniki et al., 2021)
	Reporting guideline recommendation by publishers	Weaker effect when guidelines are only mentioned, more positive changes following more active endorsement (but	(Hopewell S et al., 2012; Sims M T et al., 2018)

		mixed evidence and high diversity of interventions)	
Research material/data sharing	Publisher data sharing policies	Suggestions do not seem to be effective, mandates increase availability, sharing generally differs depending on various factors (such as disciplinary norms)	(Hamilton et al., 2023)

A majority of authors reported positive effects, meaning the investigated interventions seemed to improve reproducibility, replicability or a proxy outcome. However, these interpretations were subjective, and authors employed different thresholds, for instance, when determining which data sharing rates they viewed as a success. Due to the diversity in how studies were designed and reported, quality and strength of the evidence could not be assessed formally, with only few studies employing randomized and/or controlled designs. Still, while some results appear promising, interventions might also be affected by various barriers to reproducibility, reducing their potential effects. Studies only investigating proxy outcomes instead of using more direct measures of reproducibility or replicability also do not yet allow for sound conclusions about impact pathways towards improved reproducibility. There is evidence suggesting that single proxy outcomes are not sufficient to achieve reproducibility, even though assumed to be closely linked. For instance, even when data is shared, reproducibility is not automatically achieved (e.g., Hardwicke et al., 2018, 2021; Laurinavichyute A et al., 2022).

Read more:

- Dudda, L., Kormann, E., Kozula, M., DeVito, N. J., Klebel, T., Dewi, A. P. M., Spijker, R., Stegeman, I., Van den Eynden, V., Ross-Hellauer, T., & Leeflang, M. M. G. (2025). Open science interventions to improve reproducibility and replicability of research: A scoping review. *Royal Society Open Science*, 12(4), 242057. https://doi.org/10.1098/rsos.242057

3.2. The evidence base for Machine Learning-driven research

Authors: Simone Kopeinik, Tony Ross-Hellauer (Know Center)

We next scoped the evidence specific to Machine Learning-driven research, using a semi-structured literature review methodology (Semmelrock et al., 2025). Machine Learning methods are increasingly deeply integrated into research methods, not just in computer science but across disciplines (Dwivedi et al., 2021). Indeed, recipients of the 2024 Nobel Prizes for both chemistry and physics included ML researchers. Hence, issues regarding the reproducibility of ML raise urgent concerns about the reliability and validity of findings not only for computer scientists but for large swathes of cutting-edge scientific research across disciplines. These methods also introduce distinct concerns regarding reproducibility (e.g., non-determinism) which make them interesting to study in the context of epistemic diversity.

We found that reproducibility in Machine Learning (henceforth ML) research is hindered by both procedural and technical barriers. A major issue is lack of transparency, including incomplete reporting of code, data, and experimental details, which makes it impossible for others to repeat or validate results (Pineau et al., 2021). Where materials are shared, many ML studies do not adhere to standards for sharing datasets, source code, and configurations. In addition, ML training processes are highly sensitive to factors such as random seeds, hardware, and preprocessing choices. These conditions, along with inadequate documentation, poor experimental practices, and unclear definitions of what constitutes reproducibility, mean that even studies that appear reproducible often yield results that differ substantially when independently tested (Gundersen & Kjensmo, 2018).

Next we identified countermeasures proposed within the literature and classified them organised them according to three main topics: technology-based, procedural and awareness/education measures (see Figure 3).

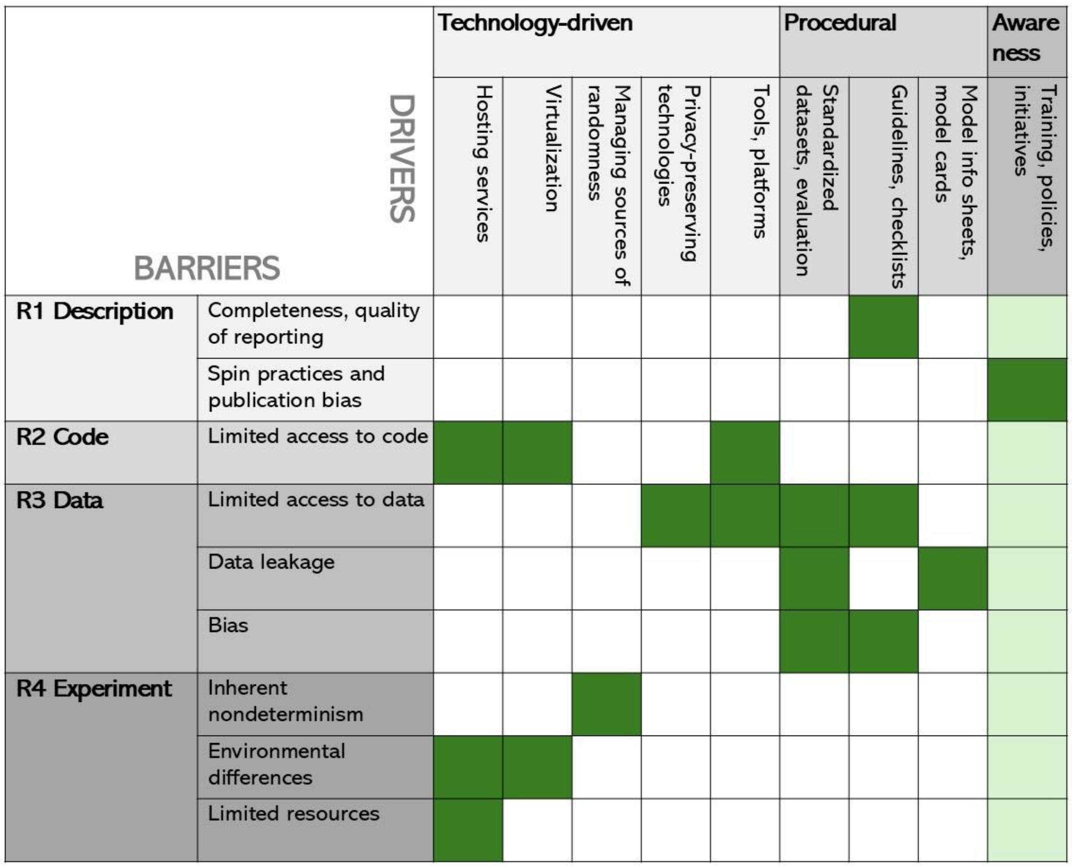

Figure 3: Barriers-Drivers Matrix. The colour indicates to what extent a barrier can be addressed with a given driver. Thus, the light-green colour of the awareness driver means that it could be used to address all barriers, but other drivers indicated with dark-green colour are more effective for addressing certain barriers.

Technology-based measures leverage technology to enhance code accessibility and reproducibility. Depending on the algorithm and the methodological approach, one or more of these actions may be necessary.

- Using controlled study environments: Virtualisation technologies (like Docker or Singularity) allow bundling the complete software environment with the code. This addresses the challenge of environmental inconsistencies that, if present, may cause a deviation in computational accuracy (Boettiger, 2015; Chirigati et al., 2016; Moreau et al., 2023).

- Using code and data hosting: Platforms such as GitHub, GitLab, or dedicated data repositories ensure permanent, accessible storage for code and data by providing an easy-to-implement solution to address barriers associated with limited access (Hartley & Olsson, 2020; Schlegel & Sattler, 2023; Soiland-Reyes et al., 2022).
- Managing randomness in ML models: Randomness forms a native part of ML model training, as it is standard practice in several stages, such as weight initialisation or data splitting. It is a source of variability that must be managed when trying to replicate model results. Thus, the creation of deterministic practices - by fixing random seeds across all relevant processes – is essential to ensure consistent results across runs, mitigating inherent nondeterminism (Ahmed & Lofstead, 2022; Ferigo et al., 2020; Pouchard et al., 2020; Raste et al., 2022).
- Data privacy solutions: Exploring privacy-preserving technologies (like federated learning or synthetic data generation) enables responsible data sharing while protecting sensitive information (Xu et al., 2021).

Procedural measures emphasise transforming research documentation and evaluation.

- Standardised Guidelines and Checklists: Implementing formal guidelines—such as reproducibility checklists—ensures researchers include all essential information in their publications (Artrith et al., 2021; Pineau et al., 2021).
- Model Transparency: Using tools like Model Cards or Model Info Sheets to offer detailed documentation about model design, data, performance, and ethical considerations helps eliminate vague descriptions (Kapoor & Narayanan, 2023; Mitchell et al., 2019).
- Standardised Data Use: Employing well-documented datasets and clearly outlining train/validation/test splits ensures fair comparisons and minimises data leakage (McDermott et al., 2021; Pushkarna et al., 2022).
- Publication Incentives: Initiatives such as Reproducibility Badges promote a culture where researchers are encouraged and rewarded for sharing code and data.

Awareness/education measures that the literature suggests contribute towards a cultural shift that complements technology-based and procedural measures.

- Training and education: Integrating training on reproducibility practices (such as version control, coding standards, and effective data management) into academic programs and professional development is essential.
- Promoting preregistration: Encouraging researchers to preregister study designs, hypotheses, and analysis plans prior to data collection minimizes publication bias and reduces the risk of selective reporting.
- Community advocacy: Actively engaging with the broader ML and scientific community reinforces the importance of reproducibility as a fundamental scientific principle, counteracting the temptation to prioritize novelty over precision.

Although each measure has been proposed as contributing to reproducibility in machine learning, the literature indicates that reproducibility in ML-driven research is best supported through a holistic approach that combines technology-based measures, procedural measures, and awareness and education initiatives. In summary, we propose that reproducibility in machine learning (ML) research will be most effective when a balanced and informed approach to the measurements described above is taken. This approach should emphasise strict compliance with reporting standards, the use of state-of-the-art technological solutions for sharing code and data, and a significant cultural shift toward greater transparency.

Read more:

- Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Haberl, A., Thalmann, S., & Kowald, D. (2025). Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Magazine*, 46(2), e70002. https://doi.org/10.1002/aaai.70002

3.3. The evidence base for qualitative research

Author: Nicki Lisa Cole (Know Center)

Next, using the integrative literature review methodology, we investigated 1) the conceptualisation of reproducibility in relation to qualitative research and 2) enablers of and barriers to the reproducibility of qualitative research. We screened 3,215 publications and included 248 in our study report (Cole et al., 2024). Reflecting the real-world manifestation of our project's conceptual framework that illuminates the relationship between knowledge production modes and the relevance and feasibility of reproducibility as a quality criterion (Ulpts & Schneider, 2025b), we found that conceptualisations of reproducibility and replicability that stem from quantitative standpoints are framed as inappropriate practices and epistemic criteria for (most) qualitative research. However, when conceptualised in alternative ways that are adapted to the epistemic conditions, aims and practices of qualitative research, they can be both applicable and appropriate. For example, it may be possible to conduct conceptual replication, using a different study design to test previous hypotheses, in a qualitative research context (Tuval-Mashiach, 2021). Methodological repeatability may be possible and allows for the transferability of findings from one study to another context or can reveal equally valid yet different findings due to a different context (Buckley et al., 2022). Another approach that can apply in some qualitative research contexts isresult reproducibility, when one reaches the same results as an original study, with similar but not necessarily the same methods (Goodman et al., 2016). Replication-in-thought enables the readers of a study to trace the design, conduct and analysis to imagine reaching the same results and conclusions (Büthe & Jacobs, 2015), and qualitative replication entails replicating one or more aspects of an original study and including a comparison with the original in reporting (Talkad Sukumar & Metoyer, 2019). It is important to note, however, that these ways of redoing serve different epistemic functions than when quantitative studies are replicated or reproduced. Rather than looking to verify a study design, methods and findings, redoing in qualitative research might aim to identify transferability of findings from one context to another, or to compare biases, prejudgments and thought processes among different researchers.

We identified several barriers to qualitative research reproducibility through our review. The primary barrier is the ontological and epistemological misalignment of reproducibility, replicability and Open Science practices with qualitative research. While in quantitative approaches, researchers often start with the assumption of a fixed, knowable truth that they can identify through particular research methods; many qualitative research approaches do not align with this. They may be constructivist or interpretive in nature and rooted in and shaped by researcher subjectivity.

Some key aspects of qualitative research feature prominently in these discussions — the context-based nature of qualitative research — always situated within a particular time and place, with a particular group of people — and the fundamental role of the researcher in shaping the data that is collected, and in interpreting it.

Other key barriers include anonymity, ethics and consent, which are important concerns when working with participants, and these are most often framed as challenges to data sharing and data reuse, specifically, which forms a barrier to qualitative reproducibility. Concerns around anonymity include the potential risk to participants of reidentification after data are shared, that anonymization processes are challenging and costly (in terms of time and staffing), and that fully deidentified data may be practically useless in terms of reuse. Ethical barriers arise when "the ethical imperative of open science" conflicts with research ethics and accountability to participants (Prosser et al., 2022), and concerns exist as to whether informed consent can be appropriately carried out or achieved regarding future users and uses of shared data.

Despite these barriers, our review also found several enablers of open qualitative research that may support forms of qualitative reproducibility. In fact, often, the very things that are framed as barriers are also framed as enablers. To enable qualitative data sharing and reuse, we found careful anonymization processes, ethical consent practices and processes, deep research context documentation, and ethical data access management suggested within the literature. We also found adaptation and flexibility of Open Science practices and norms to be key enablers. This includes adapting guidance, training, templates (like preregistration) and infrastructures to suit qualitative research designs, methods, and data.

Finally, we found that established qualitative research practices, designed to support transparency, are enablers of Open Science practices, and potentially some forms of reproducibility. Documentation, by providing context details and insight into the analytic process, can provide information necessary for another researcher, or educator, to effectively and ethically reuse qualitative data. The practice of reflexivity, and incorporating it into documentation, can help to balance out the barrier to reuse that is thought to be created by researcher subjectivity and positionality, especially when it can illuminate the interpretive process and the role of the researcher in shaping it. The establishment of rapport and trust between researchers and participants can enable active, ongoing informed consent to support ethical data sharing. This is the basis for having ethical discussions with participants about data sharing and reuse, and for involving them in decisions on these matters. Importantly, our study found that these practices are also understood to enable open research processes, like Open Methods and Open Analysis. These, in turn, support data sharing and data reuse.

To conclude, our study found that reproducibility and Open Science practices must be adapted to the aims and epistemic conditions of qualitative research for them to be applicable and feasible, and that they will not always be relevant nor feasible for all qualitative research. The pathway to reproducibility of qualitative research is therefore hedged with this caveat. However, in some cases, achieving certain forms of reproducibility or replication may be possible for qualitative research, when ethical data sharing is possible and reuse is responsibly and ethically managed, and when existing rigorous and transparent qualitative research practice are leveraged to achieve

Open Methods and Open Analysis. Indeed, such practices can be leveraged to increase the openness of research methods and analysis across diverse disciplines and epistemologies, not just within qualitative research.

Read more:

- 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. *MetaArXiv*. https://doi.org/10.31222/osf.io/n5zkw_v1

4. Enhanced understanding (3): TIER2 original studies

4.1. Forecasting the future of reproducibility

Joeri Tijdink (VUmc)

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 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 required for sharing research materials. 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.

The future of reproducibility that participants imagine:

- Has a particular research culture that prioritises quality over quantity and centres reproducibility in research practice and in training;
- Has standardised reproducibility requirements that account for methodological and epistemic diversity and standardised and shareable methods, tools and workflows;
- Incentivises reproducible, open and collaborative practices by providing recognition for them, funding them, and making them visible (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 sufficient resources are available to develop and maintain such infrastructure.

Our results generally align with previous studies that have assessed the implementation of open science practices and reproducibility, indicating a need for a culture change and training for all actors involved to achieve the desired goal.

The study also identifies several tensions between enablers and barriers perceived by diverse stakeholders. First, level of standardisation or flexibility that should be maintained in the pursuit of reproducibility. Participants identified pros and cons of a common set of standards shared between researchers and other stakeholders in diverse disciplines and contexts, versus 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 (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? All in all, our conclusions echoed our participants' view that researchers should be in leading positions to develop these. However, guidance from institutional actors would be useful in setting basic standards and expectations, linking and syncing with infrastructures and services, centring epistemic diversity, and fostering community-driven initiatives.

Finally, we call for approaches to reproducibility that focus on the full research ecosystem and lifecycle, including assessment procedures as a prime lever for initiating transformation. We hence 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.

Read more:

- Horbach, S. P. J. M., Cole, N. L., Kopeinik, S., Leitner, B., Ross-Hellauer, T., & Tijdink, J. (2026). How to get there from here? Barriers and enablers on the road towards reproducibility in research. *Journal of Trial and Error (forthcoming)*. OSF Preprint: https://doi.org/10.31222/osf.io/gx9jq_v1
- Tijdink, J., Leitner, B., Cole, N., Horbach, S., Kopeinik, S., & Ross-Hellauer, T. (2023). TIER2 D4.1 The Future(s) of Reproducibility in Research. https://osf.io/dzq9e

4.2. Auto-ethnography of reproducibility practices in TIER2

Nicki Lisa Cole (Know Center)

A particular innovation in TIER2 was our project's commitment to self-reflection upon our own reproducibility practices, enabled via an autoethnographic study in which TIER2 researchers continually reflected upon our own commitments to practicing and evaluating "radical reproducibility" and transparency in our own work (Cole & Horbach, 2025). The project sought to model best practice by making data FAIR, preregistering protocols, openly sharing software and workflows, and fostering an environment that embraces epistemic diversity. Recognising that reproducibility is interpreted differently across national, disciplinary, and epistemic cultures, the consortium undertook a structured internal reflection to better understand the challenges, benefits, and limitations of implementing reproducibility practices in an international, interdisciplinary research setting.

The Autoethnography study employed a longitudinal, 3-year autoethnographic approach, combining consortium-wide discussions at General Assemblies, pre-meeting reflective surveys, and quarterly reproducibility diaries authored by a diverse set of team members. We found that consortium members' understanding of reproducibility deepened over the project. Initial differences in definitions—linked to disciplinary and epistemic traditions—gave way to a more nuanced appreciation of its conceptual and practical complexity. Participants developed new skills in data sharing, preregistration, workflow design, and critical evaluation of research practices. Many reported "aha moments," such as recognising the importance of early workflow planning and the need for shared training within teams. Members expressed enthusiasm for implementing open and transparent workflows, pride in the project's commitment to epistemic diversity, and satisfaction with the collaborative environment. The use of Open Science Framework, co-creative tool development, and structured documentation practices were seen as successful enablers of reproducibility. The reflective process itself was often described as intellectually rewarding and professionally empowering.

From a critical standpoint, members highlighted emotional discomfort around exposing imperfect work, significant time and resource burdens, and uneven motivation and engagement across career stages. Technical and methodological obstacles—especially regarding qualitative data, evolving research questions, and uneven training and experience—acted as challenges to reproducibility. Epistemic tensions surfaced regarding the feasibility and relevance of reproducibility across qualitative and quantitative paradigms, with some members fearing over-complication while others stressed the importance of acknowledging epistemic diversity. Perceived enablers included strong role modelling by senior researchers, improved documentation practices, open-source tools, containerisation of workflows, and "slow science" approaches that prioritise depth and quality over speed and quantity. Members emphasised the need for better incentives, more consistent training, culture change within research institutions, and stronger evidence on the impact and cost-effectiveness of reproducibility interventions.

The TIER2 project hence demonstrates that pursuing radical reproducibility in a diverse, multi-partner consortium yields significant learning, skill development, and cultural transformation, but also reveals practical constraints and epistemic tensions that shape what can realistically be achieved. Despite moments of anxiety, resource limitations, and conceptual disagreement, the approach proved valuable for deepening understanding, improving practice, and amplifying reproducibility beyond the project itself. The findings reinforce calls for enhanced funding, institutional incentives, and evidence-based policy to support reproducibility across diverse research contexts.

Read more:

- Cole, N. L., & Horbach, S. (2025). TIER2 D1.4 Autoethnographic reflections on implementing radical reproducibility in the TIER2 project. https://osf.io/nhak3

5. Innovative tools and practices: The TIER2 Pilots

At the heart of TIER2 were a series of eight pilot activities to develop, implement and assess new reproducibility tools and practices relevant to researchers, funders and publishers. This chapter presents evidence from the pilots and other knowledge-generating project activities that demonstrate key impact pathways to reproducibility and which tools and practices can enable them. Full overviews of processes and outputs are available in the reports listed below. In this section, we focus on key outcomes and lessons learned.

Read more:
- Leitner, B., Tijdink, J., Kohrs, F., Bannach-Brown, A., Ulpts, S., Momeni, F., Adamidi, E., Klebel, T., Kormann, E., & Marangoni, A. (2025). TIER2 D4.3 Pilot Implementation Report. https://osf.io/7e6dy
- Adamidi, E., Vergoulis, T., Momeni, F., & Papadopoulou, E. (2025). TIER2 D5.1 Tools and practices for researchers. https://osf.io/5nqh6
- Klebel, T., & Lister, A. (2025). TIER2 D5.2 Tools and practices for publishers. https://osf.io/s7gjv
- Adamidi, E., Vergoulis, T., Tijdink, J., Leitner, B., Stavropoulos, P., Amodeo, S., & Papageorgiou, H. (2025). TIER2 D5.3 Tools and practices for funders. https://osf.io/pfjth

5.1. Pilot 1: Decision Aid: Relevance and Feasibility of Reproducibility

Author: Jesper Schneider (Aarhus University)

This Pilot built upon the conceptual work detailed in Section 2 to investigate the operationalisation of a guided decision tool that could help indicate whether reproducibility was relevant and, if so, to what extent it was feasible in the given context. None of this was pre-planned and written into the TIER2 application. The idea emerged at the end of the work in Task 3.1 and as such the resources available for development were limited. In short, the prototype consisted of two components: Relevance and feasibility. The basic idea in the relevance component was to allow the user to map the intended epistemic function (purpose) to the actual parts of the research which should be varied or kept the same (the actual practices).

In practice, it was found that optimal use presupposed substantial knowledge of methodology and epistemology, as well as detailed familiarity with the specific research under consideration. In other words, the complexity was high—indeed too high—given that the intended stakeholders included funders, i.e., external users tasked with evaluating a piece of research (e.g., a grant application). The aim of acknowledging and operationalizing epistemic diversity inevitably led to a degree of complexity that required an unfeasible level of expertise from the intended users. Therefore, while we think that it is still an insightful and useful analytical aid, it might be a too fine-grained unit of analysis to operationalize into a workable tool. This revision of the prototype was never fully developed and therefore never reached the stage of a pilot-testing. Resources did not allow us to take this further.

5.2. Pilot 2: Reproducibility Management Plans

Author: Stefania Amodeo (OpenAIRE)

The Reproducibility Management Plan (RMP) pilot aimed to develop both a conceptual framework and technical implementation to enable systematic planning of reproducibility across the research lifecycle. The technical implementation leveraged the ARGOS platform[1] to produce machine-actionable plans aligned with the DMP Common Standard, integrated with persistent identifiers and FAIRsharing APIs. Through extensive co-creation with 89 stakeholders across 12 European and 3 non-European countries, the pilot extended familiar Data Management Plan practices to comprehensively address reproducibility needs. Methods included focus groups with scientists and reproducibility professionals to identify planning-stage questions, policy workshops with funders, monthly ARGOS community calls for usability feedback, and real-world testing with CHIST-ERA projects[2] completing RMPs.

Three key enablers of reproducibility emerged from the pilot: integration with existing DMP workflows, structured prompts that guide researchers through systematic thinking without requiring specialized training, and machine-actionable formats that enable systematic monitoring by funders rather than relying on anecdotal assessment. However, challenges also emerged, including substantial time investment (45–68 minutes to complete an RMP), variability in output quality indicating the need for better guidance, and critical dependence on institutional support infrastructure. The pilot demonstrated innovation by establishing a framework that shifts reproducibility from post-hoc verification to proactive planning from project inception. While challenges remain, this shift from reactive reproducibility verification to proactive reproducibility planning represents a fundamental change in how we approach research quality.

5.3. Pilot 3: Reproducible Workflows

Authors: Eleni Adamidi and Thanasis Vergoulis (ARC)

Pilot 3 focused on enabling reproducible computational research through the development of the SCHEMA api and SCHEMA lab,[3] an open-source virtual laboratory supporting containerized task execution, workflow execution and management, and computational experiment creation. The pilot targets researchers in the life and computer sciences, where pipeline complexity and data scale challenge transparency and reuse. By combining containerization,[4] workflows, and structured metadata, the pilot promotes the creation of computational experiments that are shareable, and thereby, embedding reproducibility into the research process itself rather than treating it as a post-hoc activity. The development process followed a co-creation approach, involving iterative engagement between life science and computer science researchers. The stakeholders contributed to the requirements collection, user interface design and computational workflow testing to ensure that the tools reflect real-world research needs and usability expectations.

[1] https://argos.openaire.eu/portal/

[2] CHIST-ERA is a coordination and co-operation activity of national and regional research funding organisations mainly in Europe and is supported by the European Union. See: https://www.chistera.eu/

[3] Main website: https://schema.athenarc.gr/about/, API: https://github.com/athenarc/schema-api, SCHEMA lab: https://github.com/athenarc/schema

[4] Containerisation packages software, data, and dependencies into isolated, reproducible environments that ensure consistent results across different systems in data management and research.

The pilot demonstrates that reproducibility is strengthened when computational experiments are expressed as structured workflows that capture the full computational context such as software, parameters, data, and provenance, allowing results to be visible and shareable. These workflows and experiments including workflows act as clear pathways to reproducibility by transforming ad hoc analyses into machine-readable research objects that support reuse. Key enablers include the integration of metadata standards, execution environments, and traceable experiment packaging. Persistent challenges include limited awareness of technologies such as RO-Crates and potential uneven technical capacity. Overall, the pilot confirms that promoting workflow-based experimentation provides a sustainable and scalable foundation for reproducible science.

Read more:

- Adamidi, E., Deligiannis, P., Foutris, N., & Vergoulis, T. (2025b). A Virtual Laboratory for Managing Computational Experiments. Proceedings of the 37th International Conference on Scalable Scientific Data Management, 1–6. https://doi.org/10.1145/3733723.3733743

5.4. Pilot 4: Reproducibility Checklists for Computational Social Science Research

Author: Fakhri Momeni (GESIS)

Pilot 4 provides empirical evidence that structured, checklist-supported workflows can improve reproducibility in computational social science. Within TIER2, the Methods Hub platform (https://methodshub.gesis.org/) served as the implementation environment where a reproducibility checklist was developed, validated, and integrated. Two surveys and an experimental study demonstrated how concise, actionable guidance supports reproducible research.

Survey results (N = 180) revealed that researchers often value reproducibility but lack clear, practical instructions. Based on this input, a simplified checklist was created, focusing on essential elements (data, code, environment, and sharing) and embedded in the Methods Hub's metadata fields.

Experimental evaluation confirmed its effectiveness: reproducibility success increased to 84 % on Methods Hub compared with 72 % on external repositories, with fewer repository-related errors and reduced reliance on external help. These findings show that usability and simplicity are crucial for adoption.

Pilot 4 thus evidences a clear impact pathway:

Guidance → Simplified Checklist → Platform Integration → Improved Reproducibility Outcomes.

This demonstrates that embedding lightweight, user-oriented reproducibility requirements into research infrastructures directly strengthens reproducibility in practice.

Read more:

- Momeni, F., Khan, M. T., Kiesel, J., & Ross-Hellauer, T. (2025). Checklists for Computational Reproducibility in Social Sciences: Insights from Literature and Survey

Evaluation. Proceedings of the 3rd ACM Conference on Reproducibility and Replicability, 179–191. https://doi.org/10.1145/3736731.3746161

5.5. Pilot 5: Reproducibility Promotion Plans for Funders

Author: Joeri Tijdink (VUmc)

Funders hold an integral role in influencing the reproducibility, openness, and transparency of research due to the available means they possess to shape research practices and promote reproducibility. However, funders are overburdened in their work and may not have the necessary tools to drive policy changes. Pilot 5 co-created the Reproducibility Promotion Plan for Funders (RPP). The RPP serves as a policy template with recommendations across three key areas of funding work: policy and definitions, evaluation and monitoring, and incentives. The RPP provides actionable recommendations alongside guidance and best practice examples which funders and funding institutions can adapt to meet their specific needs.

The RPP was piloted over a six-month period in both an international and a national funding organization. The pilots demonstrated that reproducibility can be incorporated into funding organizations in various ways and is applicable across a range of contexts, regardless of epistemic orientation or disciplinary focus. Participants reported that the RPP was both applicable and usable, noting its customizability to meet their specific needs. However, they also highlighted several barriers unrelated to the RPP itself, identifying bureaucratic processes, time constraints, and financial limitations as major obstacles to implementing policy changes.

Read more:

- Tijdink, J., Leitner, B., Kohrs, F., & Bannach-Brown, A. (2025). Reproducibility Promotion Plans for Funders. https://osf.io/49gfw

5.6. Pilot 6: Reproducibility Monitoring Dashboard

Author: Haris Papageorgiou (ARC)

The Reproducibility Monitoring Dashboard serves as a comprehensive platform designed to provide stakeholders—including funding agencies and research organizations—with systematic tracking and monitoring capabilities for evaluating the adoption and implementation of reproducible research practices. The primary purpose of this pilot initiative was to enhance transparency in research by establishing a systematic framework for monitoring reproducibility metrics, thereby supporting both evidence-based policy development and compliance assessment. This overarching goal was operationalised through four key objectives that address critical research questions in the domain of research reproducibility. First, the pilot developed and validated a robust and explainable suite of tools capable of tracking major research artifacts, including datasets, software, and computational resources. Second, it quantified and estimated reusability indicators across different artifact types, establishing standardized metrics for assessing research output sustainability. Third, the pilot developed reliable proxies for reproducibility and replicability that can alleviate the evaluative burden on funding agencies and research-performing organizations (RPOs) while simultaneously providing evidence-based insights into the impact of their implemented policies. Finally, the pilot designed and implemented an interactive dashboard infrastructure that enables funding agencies and RPOs to efficiently and effectively track and monitor the reusability of research artifacts—including datasets, software, tools, and systems—generated within funded projects.

Insights from the pilot highlighted the value of coordinated, multi-level engagement in advancing reproducible research. At the funding and institutional levels, systematic monitoring supported reflection on policy implementation, portfolio performance, and areas where additional support or alignment were needed. Shared use of the dashboard infrastructure also pointed to opportunities for collaboration between funding bodies and research-performing organisations, particularly in expanding coverage across disciplines and accommodating field-specific artefacts and metrics. At the researcher level, the dashboard's potential as a discovery and learning tool underscored how visibility of well-documented, reusable outputs could reinforce reproducibility as a normative practice. Similarly, its relevance to scholarly publishing suggested how earlier signals about documentation quality could help strengthen reproducibility expectations at the point of dissemination. Taken together, the pilot illustrated how shared infrastructure, transparent metrics, and cross-stakeholder alignment can help address persistent barriers to reproducibility while reinforcing the conditions that enabled it. (Stavropoulos et al., 2023)

Read more:

- Stavropoulos, P., Lyris, I., Manola, N., Grypari, I., & Papageorgiou, H. (2023). Empowering Knowledge Discovery from Scientific Literature: A novel approach to Research Artifact Analysis. In L. Tan, D. Milajevs, G. Chauhan, J. Gwinnup, & E. Rippeth (Eds), Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023) (pp. 37–53). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.nlposs-1.5

5.7. Pilot 7: Editorial Workflows to Increase Data Sharing

Author: Thomas Klebel (Know Center)

A key aspect of (computational) reproducibility is availability of data. However, sharing of research data is still not the norm across disciplines. The workflow and email template developed in Pilot 7 aimed to provide a low-effort approach for publishers to nudge researchers towards sharing their data for journals operating under a "share upon request" policy. Preliminary findings from a randomised controlled trial to evaluate the intervention's efficacy show that the rate of authors sharing their data in a trusted repository is low in the analysed journals. Our analysis of the intervention's efficacy shows that the workflow has no substantive effect on the rate of researchers sharing their data in a trusted repository. The trial provided weak evidence for a small reduction in researchers opting to share data on request, stating that data were unavailable, and other forms of data sharing statements. Further analyses will explore the changes to the DAS texts undertaken by the authors, to disentangle potential effects of the intervention from other editorial actions undertaken by the publisher aimed at improving data sharing practices.

In light of other approaches to changing researcher behaviour on sharing data (see for example Pilot 8), our results are not unexpected: to obtain substantial changes towards broader availability of research data and subsequently reproducibility of results, publishers would need to implement stricter policies mandating sharing of research data, along with adequate monitoring and enforcement. Further information on the intervention and its implementation can be found in TIER2 deliverables D4.3 and D5.2.

Read more:

- Klebel, T., Kormann, E., Van den Akker, O., & Ross-Hellauer, T. (2024). Impact of sharing information on how to share data with authors submitting manuscripts to journals to improve Data Availability Statements: Protocol for a randomised controlled trial. https://osf.io/d9v47

5.8. Pilot 8: An Editorial Reference Handbook for Reproducibility and FAIRness

Author: Allyson Lister (Oxford University)

The Editorial Reference Handbook[5] was developed collaboratively, by 35 participants representing 19 journals and 11 publishers, through a series of eight online workshops conducted in 2024. The Handbook integrates structured checks, narrative guidance, and visual workflows to bridge the gap between policy and editorial practice. It can assist journals and publishers in two primary ways: (i) for those without internal guidance to enforce an open research policy, it provides a workflow for assessing and improving the openness of individual manuscripts; and (ii) for those with existing guidance, it offers principles that can be used to validate and enhance current practices. It provides a model for embedding good research practices and FAIR principles into the scientific publication process, while also exemplifying the broader cultural shift toward open and responsible practices within scholarly publishing. By requiring clear identification, deposition and formatting of digital research objects, the Handbook supports manuscript authors in the publication of well described and discoverable DOs, a key part of FAIR. By offering a shared, operational resource grounded in a consensus set of small but practical checks aligned with journal roles and workflows, the Handbook addresses a critical gap and supports scalable adoption. By implementing the concepts within the Handbook, publishers and journals are improving the FAIRness of their publishing practices and of the digital research objects described in conformant manuscripts.

The Handbook was also piloted through a structured intervention across 190 manuscripts, with outcomes evaluated using both surveys and performance metrics. All participants agreed that the Handbook, with its small but practical set of checks, is a valuable aide to good research practice that they intend to continue implementing. Their feedback showed that the Handbook elements are feasible to implement in real-world editorial contexts. Differences in publisher methodologies and implementations—including variations in repository checks, policies, and staff roles—might initially appear to present a barrier to the Handbook's aim of aligning checks across journals and publishers. However, the experiences of both intervention participants and positive controls demonstrate that the Handbook is sufficiently rigorous to be educational and practical, while also retaining the flexibility necessary for adoption across diverse journal contexts, tailored to local readiness and priorities.

[5] Available at: https://publishers.fairassist.org/

Read more:

- Taylor-Grant, R., Cannon, M., Lister, A., & Sansone, S.-A. (2025). Making reproducibility a reality by 2035? Enabling publisher collaboration for enhanced data policy enforcement. International Journal of Digital Curation, 19(1). https://doi.org/10.2218/ijdc.v19i1.1064
- Lister, A., Taylor-Grant, R., Cannon, M., Ahmed, R., Alfarano, G., Begum, R., Bright, J., Cadwallader, L., Cranston, M., Dunkley, L., Edmunds, S. C., Flammer, P., Hunter, C., Hyde, A., Klebel, T., Leary, A., MacCallum, C., McKenna, S., McNeice, K., … Sansone, S.-A. (2025). *Supporting FAIR Practices In Scholarly Publishing with the Editorial Reference Handbook* (No. 9vujt_v2). MetaArXiv. https://doi.org/10.17605/OSF.IO/FB9QW

6. Increased capacity

A core principle of TIER2 was that action to increase reproducibility must be targeted holistically to boost capacity at all levels. Our consortium's expertise in Research Integrity, Open Science, Scholarly Communication, and disciplinary Research Infrastructures, as well as our close contacts to research communities and networks of publishers and funders, meant that we were uniquely placed to play a key role in boosting capacity. In this section we give an overview of TIER2's many capacity-building activities and outputs.

6.1. TIER2 Reproducibility Hub

Authors: Barbara Leitner and Joeri Tijdink (VUmc)

The TIER2 Reproducibility Hub (ReproHub)[6] serves researchers, funders, publishers, institutions, and other stakeholders, providing a centralized resource to enhance reproducibility. To ensure sustainability, the Hub is actually a project page hosted as a sub-site of the *Embassy of Good Science*.[7] It integrates curated content from TIER2 as well as related initiatives such as the TIER2 sister projects iRISE and OSIRIS, including checklists, training modules, inventories of tools, and domain-specific guidance. Beyond serving as a registry, the Hub promotes community building by offering guidance and materials to support the establishment of Reproducibility Networks and facilitate the adoption of best practices. The Reproducibility Hub describes specific project outcomes, addresses critical gaps in current knowledge, provides a platform for evidence-based practices, gives an overview of achievements of our community, and ensures the long-term promotion of reproducibility and research integrity within the global research and innovation ecosystem. All this is structured under intuitive modules: Communities and collaborators; Future of Reproducibility; New Tools and Services for Reproducibility; What is reproducibility?; and What is the evidence for reproducibility in different epistemic contexts?

6.2. Co-creation and stakeholder engagement

Authors: Alexandra Bannach Brown (Charite)

A range of co-creation formats was employed to actively engage the TIER2 stakeholder groups at various stages of the development, implementation, and evaluation of new reproducibility-related tools and practices, emphasising stakeholder engagement and collaboration. Co-creation was initiated and managed following stakeholder mapping, the effectiveness of engagement strategies, and facilitation methods. It encompassed multiple avenues to stakeholder engagement, including Reproducibility Cafés, Co-creation workshops, Validation workshops, Webinars, ReproHacks and Tutorials. At the network level, events such as the "Building Bridges: Strengthening Reproducibility & Open Science Networks Across Europe" illustrated the value of combining formal discussion with open exchange. The meeting successfully bridged communities of National Open Access Desks (NOADs) and Reproducibility Networks (RNs), allowing participants to identify overlapping goals and efforts as well as discuss new opportunities for collaboration. The TIER2 Milestone Report "Self-reflection on Co-creation activities in TIER2"

[6] https://embassy.science/wiki/Initiative:286109fc-03cb-4a08-bd45-c0276eae3079
[7] https://embassy.science/wiki

(Kohrs & Bannach-Brown, 2025) summarises and shares our experiences and processes, reflecting on lessons learned regarding inclusivity, cross-disciplinary collaboration, and mutual learning. In particular, the report highlights the way in which iterative design processes used in our pilots enhanced community ownership and built cross-sector trust, as well as our use of Diversity, Equity, Inclusion, and Accessibility (DEIA) resources and reflexive facilitation practices to enhance collective awareness and sensitivity of representation, accessibility, and equitable participation, factors often under-addressed in reproducibility and Open Science policy discussions

.

6.3. New Reproducibility Networks in Widening Countries

Authors: Alexandra Bannach Brown (Charite)

To further expand and institutionalise reproducibility communities and to strengthen reproducibility efforts, TIER2 issued an open call for three Reproducibility Networks (RNs) in Widening Participation countries. RNs serve as national hubs advocating for rigorous, open, and high-quality research (UK Reproducibility Network Steering Committee, 2021). RNs facilitate interdisciplinary collaborations and discussions among scientists and other stakeholders, such as funders and publishers. Further, they provide training and infrastructure to build capacity. The widespread presence of RNs is crucial, as scientific communities across contexts (e.g., disciplinary, demographic, and geographic) face different challenges and barriers, and are at different stages of readiness to implement reproducible research practices. Following our initial open call in 2023, two new RNs in Ukraine and Georgia were successfully established. To complement the existing efforts in Ukraine and Georgia, a second open call was issued for the third TIER2 award. A consortium in Serbia was awarded in 2024. All supporting documentation on the RN awards including details how to issue an open call to and on the evaluation process are publicly available via OSF (Kohrs & Bannach-Brown, 2025). The success of these small awards in fostering these new Reproducibility Networks demonstrates the grass-roots demand for such support.

6.4. Training modules

Author: Stefania Amodeo (OpenAIRE)

The TIER2 Reproducibility Training modules[8] are free courses designed for researchers, publishers, and practitioners committed to enhancing research integrity. They combine theoretical knowledge with practical guidance, covering topics such as principles of reproducibility, open science practices, methodological and epistemological considerations, operational checks, and tools to enhance transparency, reliability, and trustworthiness across different research contexts. The modules are hosted on OpenPlato, the modular learning platform developed by OpenAIRE to support training in Open Science, Research Data Management, and related topics across Europe. The full list of modules are: Introduction to Reproducibility; Understanding Epistemic Diversity; Tools and Best Practices for Reproducibility; Implementing Reproducibility in Research; Reproducibility primer for funders; Reproducibility primer for publishers; Reproducibility primer for qualitative research; and Reproducibility primer for AI-driven research.

[8] Available via OpenPlato at : https://openplato.eu/enrol/index.php?id=543

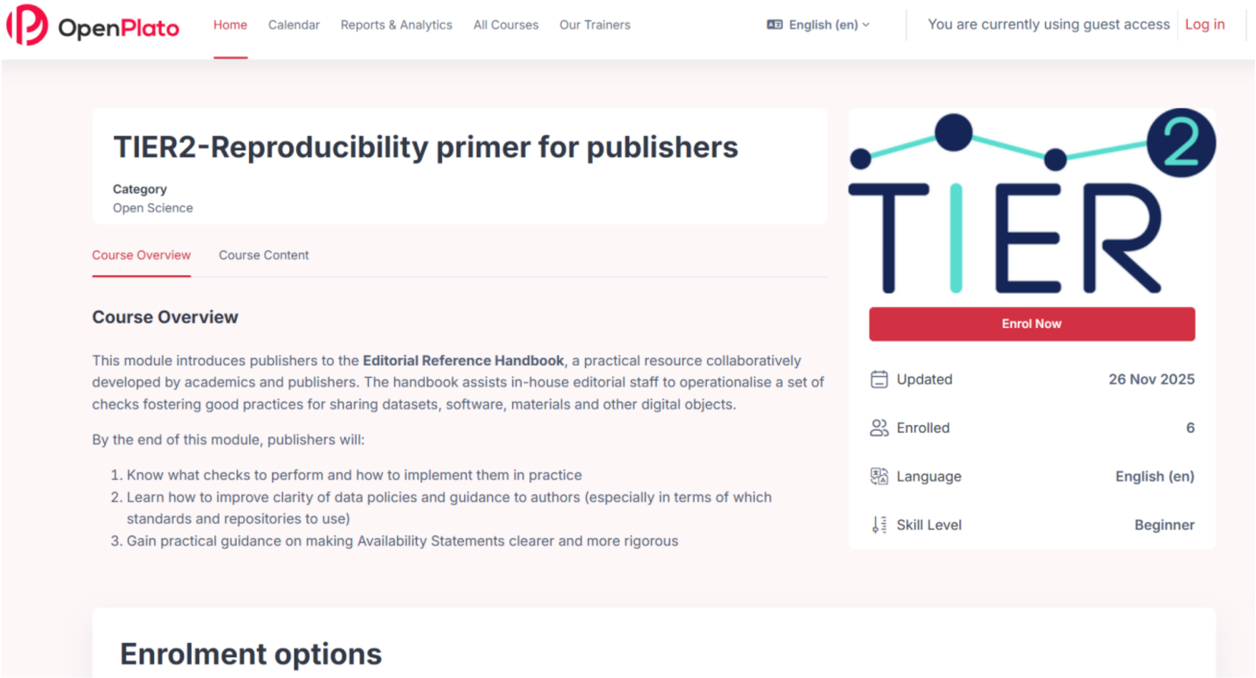

Figure 4.. Screenshot of example module from the TIER2 Training Modules

6.5. The MERRI Collaboration

Author: Tony Ross-Hellauer (Know Center)

From its outset TIER2 has collaborated closely with our sister projects iRISE[9] and OSIRIS[10] (funded under the same EC call). To continue this collaboration beyond the lifetimes of our projects, members of all three consortia have founded The Meta-Research for Research Improvement (MERRI) Collaboration.[11] MERRI is a community of practice that responds to the growing use of meta-research, or research on research, to investigate and strengthen the research ecosystem. It was established to facilitate systemic, cross-disciplinary and cross-stakeholder cooperation aimed at building a comprehensive evidence base on interventions designed to improve research transparency, reproducibility, culture and practice, including assessments of their effectiveness, potential harms and generalisability. MERRI provides an inclusive and supportive environment that values diverse perspectives and experiences, bringing together meta-researchers at all career stages and across disciplines to advance rigorous and transparent research within a positive research culture. The MERRI Collaboration's priorities include:

- Connecting Researchers in meta-research: Fostering collaboration by bringing individuals together to develop robust meta-research projects with high potential for impact.
- Mentorship: Providing structured mentorship to enhance the capabilities and integration of those new to meta-research.
- Education: Offering guided opportunities for meta-research projects, leveraging the expertise and guidance of experienced members of the meta-research community.
- Stakeholder Engagement: Engaging effectively with stakeholders, ensuring that our research efforts align with and are responsive to the broader scientific community's needs.

9 https://irise-project.eu/
10 https://osiris4r.eu/
11 https://merricollaboration.github.io/

Meta-research as a field is growing quickly, and grassroots communities such as MERRI will help develop the field and empower the next generation.

7. Policy Roadmap (1): Synthesis of lessons learned regarding key impact pathways to reproducibility

Author: Tony Ross-Hellauer (Know Center)

This section synthesises lessons learned across the evidence base, conceptual work, pilot interventions, and capacity-building activities of TIER2 to articulate validated impact pathways towards improved reproducibility. These key lessons are the results of continuous cross-consortium dialogue, as well as workshops and other feedback-gathering activities with stakeholders that we have held at the end of the project to include different stakeholder perspectives. Taken together, our findings suggest that reproducibility emerges through the complex interaction of epistemic, technical, cultural, and institutional factors. Effective measures must align these dimensions while remaining sensitive to disciplinary diversity and practical constraints. Below we describe eight key conclusions that we have drawn from the workshops and the other feedback gathering activities.

7.1. Reproducibility as inherently contextual

A central insight from TIER2 is that reproducibility is inherently contextual. Our Knowledge Production Mode (KPM) framework (Sec 2., cf. Ulpts & Schneider, 2025b) demonstrates that the relevance and feasibility of reproducibility vary across epistemic traditions, research goals, and material conditions. Treating reproducibility as a universal criterion of research quality hence risks misalignment with legitimate (other) forms of inquiry, particularly in qualitative, interpretive, or exploratory research. Evidence from our integrative review of qualitative research (Sec. 3.3., cf. Cole et al., 2024) showed that conventional notions of reproducibility and replication are often considered epistemically inappropriate in such contexts, but that adapted forms, such as conceptual replication, methodological repeatability, or replication-in-thought, can serve alternative epistemic functions including transferability, reflexivity, and transparency. These findings suggest that impact pathways will be strengthened when policies and tools make explicit which epistemic functions reproducibility is intended to serve, where its feasibility in terms of constraints of resources and other factors is accounted for, and when alternative quality criteria are recognised where reproducibility is not relevant or feasible.

7.2. Cultural change and capacity building

A key theme to emerge from our Future Studies investigation (Section X) was the importance of cultural change and capacity building as necessary complements to technical and procedural measures. These findings underlined the role of norms, values, shared definitions, and incentives within research cultures in strongly shaping engagement with reproducibility practices. Through our capacity-building activities, including creation of training modules, the Reproducibility Hub, and the establishment of Reproducibility Networks, TIER2 has addressed skills gaps and supported the development of shared understanding across diverse communities. Our pilots on Reproducibility Promotion Plans for funders (Sec. 5.6) and the Editorial Handbook for publishers (Sec. 5.8) illustrate how organisations can embed reproducibility considerations across their workflows, albeit while revealing structural constraints such as bureaucracy and limited resources.

7.3. The essential role of research communities

Recognising the inherently contextual nature of reproducibility, and the need for culture change, emphasises the importance of research communities in setting standards for reproducibility practices appropriate to their own Knowledge Production Modes. This was a key finding from our Future Studies investigation (Sec 4.1, cf. Horbach et al., 2025; Tijdink, 2023), where our participants emphasised that researchers should be in leading positions to develop and implement reproducibility standards, adequately supported by institutional actors in setting basic standards and expectations, linking and syncing with trustworthy and reliable infrastructures and services, centring epistemic diversity, and fostering community-driven initiatives. On the latter point, the success of our awards for the creation of new Reproducibility Networks in three Widening Participation countries (Sec. 6.3) demonstrates the demand from grassroots communities for support to enable such bottom-up organisations. In addition, the success of our co-creation activities across the project demonstrates the ways in which stakeholder engagement can enhance legitimacy, trust, and adoption of new tools and practices, indicating the need to include research communities to ensure sustained impact. Finally, the formation of dedicated stakeholder communities, bringing together funders, publishers, and other key actors, stimulated co-creation and generated outcomes that create strong and sustainable conditions for implementation.

7.4. The impact of early-stage planning

Across multiple strands of our work, we found that early-stage planning and proactive workflow design are essential to enable reproducibility and have positive downstream effects. Our pilots for Reproducibility Management Plans (Sec 5.2) and Checklists for Computational Social Science (Sec 5.4) illustrated how structured planning at project inception can shift reproducibility from a post-hoc compliance exercise to a proactive component of research design. Our evidence synthesis (Sec 3., cf. Dudda et al., 2025) further indicated that interventions targeting proxy outcomes, such as data sharing mandates or reporting guidelines, produced limited improvements when they were not integrated into earlier stages of research practice. Finally, our autoethnographic study (Sec 4.2, cf. Cole & Horbach, 2025) showed that early workflow planning reduced downstream uncertainty and emotional burden, while late-stage documentation was often experienced as costly and stressful. These findings collectively point to planning and design phases as critical leverage points within reproducibility impact pathways.

7.5. The importance of intuitive infrastructures and associated support

Usable, integrated, and adequately supported infrastructures also emerged as key mediators of impact throughout our Pilots. Our pilot on workflow-based approaches to computational research (Sec 5.3) enabled reproducibility by capturing software, data, parameters, and provenance in structured, shareable formats, thereby enabling semi-automated curation of reusable research objects. Our Computational Social Science Checklist (Sec 5.4) showed that simplified, platform-integrated guidance significantly improved reproducibility outcomes compared to external or optional requirements. The Reproducibility Monitoring Dashboard (Sec 5.6) further demonstrated how machine-actionable indicators and explainable metrics could support systematic monitoring and assessment to reduce costs of evaluation for funders or institutions. The Editorial Reference Handbook (Sec 5.8) shows also the importance to have underpinning infrastructure services, such as FAIRsharing, which are essential to operationalise the checks, and educate the users on how to implement the checks. However, multiple pilots highlighted persistent constraints, including uneven technical capacity, limited awareness, and dependence on institutional support, underscoring that infrastructure alone is insufficient without accompanying resources and guidance.

7.6. Mandates and structured checks may be necessary at reporting stage

Given available evidence, it seems likely that mandates and structured checks are necessary to change data sharing practices. A key finding from our evidence synthesis is that interventions during publishing workflows regarding data-sharing and/or use of reporting guidelines seem to show only weak effects where these are only recommended. Rather, stronger mandates are often necessary to motivate action. These findings from the literature are supported by our own findings from our pilot on data-sharing (Sec 5.7), where a randomised trial intervention which provided authors with motivational or instructional cues on why data-sharing is important, or how it is to be achieved, showed little effect upon rates of data-sharing. In addition, our pilot to create and implement an Editorial Reference Handbook for checks on reproducibility-related publishing elements (Sec 5.8) showed the value in publishers implementing standardised checks. Future research should further investigate how mandating improves the uptake of these practices and their direct effects upon reproducibility.

7.7. The need for meta-research

Our evidence synthesis revealed a striking lack of empirical evidence regarding the efficacy of reproducibility-related interventions. Across disciplines, most available studies focused on narrow proxy indicators. As said in the conclusion to our Deliverable report on the Pilot implementation and assessment (Leitner et al., 2025), a key strength of our Pilots was “that they were not simply conceptual exercises: most of them were tested in real contexts, involving actual research teams, journal editors, institutional offices, and funding organisations. This allowed the project to observe not only whether ideas were theoretically promising, but also how they fared when confronted with the tempo, constraints and habits of everyday research work.” Our project can be seen therefore to have highlighted the need for systematic meta-research as a critical impact pathway in its own right, necessary to move reproducibility policy beyond normative aspiration towards empirically grounded, adaptive, and context-sensitive governance. The field of meta-research is quickly growing, and our project reveals the need for more systematic evidence and support from funders, publishers and other key stakeholders for the kinds of practice-grounded meta-research conducted by TIER2.

7.8. Reflections on gains and savings

Across our Pilots, a recurring theme was the potential for system-level gains through enhanced planning to reduce downstream inefficiencies. Additionally, gains also emerged through improved documentation and packaging of research artefacts, with positive downstream impacts upon reusability, as well as supplementary benefits including methodological training. At the same time, though, the pilots made clear that these gains were contingent on usability, proportionality, and institutional support. Crucially, the pilots showed that costs and benefits were unevenly distributed across stakeholders and over time, with researchers often bearing the upfront costs of planning and documentation, lack of resources and knowledge, while many of the downstream savings accrued to funders, institutions, publishers, and future users of research outputs.

These points underline the fact that that commitments to reproducibility are not cost-neutral in terms of time and attention. For researchers, early-stage planning, learning to use available infrastructure, and producing high-quality documentation often require additional time or, at minimum, a reallocation of effort to earlier stages of the research process. Similarly, more thorough checks by publishers, while technically feasible, will require substantial and sustained resourcing that may compete with other priorities. These are therefore not simply questions of technical capacity, but of how limited time and resources are allocated across the research ecosystem. Strengthening reproducibility may entail doing less of other activities or even accepting slower publication and fewer studies overall. Recognising these opportunity costs reinforces the need for incentives, support structures, and realistic expectations that align reproducibility goals with how time and resources are invested across stakeholders.

7.9.Conclusion

Taken together, the TIER2 findings show that reproducibility is not achieved through individual tools or mandates alone, but through combinations of context-sensitive solutions, early-stage enablement, usable infrastructures, cultural and capacity-building measures, and cross-stakeholder alignment. We therefore emphasise the need for adaptive, pluralistic, and evidence-informed policymaking approaches that evolve alongside research practices and respect epistemic diversity. In our next section, we build on these findings to introduce key recommendations for future policy.

8. Policy Roadmap (2): Stakeholder co-created recommendations

Authors: Thomas Klebel (Know Center), Joeri Tijdink (VUmc), Alexandra Bannach-Brown (Charite), Tony Ross-Hellauer (Know Center)

TIER2 has prepared the following evidence-informed, collaboratively designed recommendations for actions to foster reproducibility of research. These recommendations were iteratively prepared through an inter-consortium workshop (with members of TIER2, OSIRIS and iRISE), intra-consortium discussion amongst TIER2 members, and external stakeholder feedback and validation (through in-person presentation and discussion, and in-document collaboration). They build on the empirical evidence gathered within TIER2 (described in the previous sections of this report), as well as inputs from a wide range of stakeholders. Selection of the final set of recommendations was conducted via a multi-step process. We first collected a long list of potential recommendations (v1) via workshops and requests within TIER2 and the TIER2 advisory board. Next, we collected structured feedback from colleagues within TIER2, as well as the TIER2 advisory board members, on the relevance and feasibility of each suggested recommendation, along with further conceptual feedback. From this feedback, we collated the revised list (v2), aiming to capture recommendations that are relevant, actionable, and fill gaps in existing efforts to improve reproducibility across the research ecosystem. We received feedback on the revised version (v2) from colleagues across the TIER2 consortium, seven TIER2 advisory board members, as well as from representatives across our stakeholder communities among publishers and funders. The feedback was carefully integrated into the final version of the TIER2 recommendations. Throughout this process, our priorities were to identify gaps in existing policy and provide concrete advice, and to avoid making general recommendations that were either vague of that restated points already well established.

Two conceptual approaches to understanding reproducibility were central to TIER2 and directly informed the creation of the recommendations: the importance of epistemic diversity, and the distinction between redoing and enabling redoing:

- Respecting epistemic diversity is important, because relevance and feasibility of reproducibility differ substantially between different modes of producing knowledge (see section 2.3). We therefore strove to develop recommendations that respect different ways of generating knowledge and leave room for implementing the recommendations in ways appropriate to diverse communities.
- The distinction between *redoing* research and *enabling* others to redo it (see section 2.2) clarifies what reproducibility practices are trying to achieve. Being mindful of this distinction enables us to be more alert to when and how individual reproducibility practices are relevant and/or feasible. Practices that enable redoing – such as transparent workflows and FAIR outputs – also create conditions for scrutiny and reuse, whether or not reproduction is actually undertaken. Our recommendations therefore address both redoing and enabling redoing of research.

The recommendations are designed to support the work of researchers, institutions, funders, publishers and other stakeholders. We emphasise that our recommendations are aimed at all types of funders – including public, governmental, not-for profit, philanthropic and commercial. Likewise, there is substantial diversity in publishers (from large commercial to small community-owned publishers), and our recommendations aim to be applicable to all. We present recommendations according to four themes. An overview is presented below in Table X

Table 3. *Overview of TIER2 recommendations*

Theme	No.	Recommendation
Infrastructure, standards and community	R1.1	Governments, institutions, and all types of funders should provide sustainable support for open infrastructures
	R1.2	Research communities should improve and expand standards and guidelines for data re-use
	R1.3	Funders, publishers and institutions should require Globally Unique, Persistent and Resolvable Identifiers (GUPRIs) and open metadata for research objects and entities
Incentives and policy	R2.1	Funders, publishers and meta-researchers should develop and implement responsible metrics to enable monitoring of reproducibility practices
	R2.2	Funders should actively support and incentivise replication studies across all funding streams
Training and skills	R3.1	Institutions, funders, and publishers should build sustainable support networks and training ecosystems for reproducibility
	R3.2	Institutions should strengthen leadership engagement and provide training to research leaders to foster reproducibility practices
	R3.3	Publishers should enhance journal capacity and infrastructure for checking and managing digital research objects
Strengthening the evidence-base	R4.1	Meta-researchers, supported by funders, publishers and institutions, should investigate the efficacy of reproducibility interventions
	R4.2	Meta-researchers, supported by funders and publishers, should investigate the costs and benefits of reproducibility interventions
	R4.3	Funders and publishers should enable meta-research regarding funding and publishing workflows through streamlined processes for collaboration and data-access

8.1. Infrastructure, standards and community

Robust, sustainable infrastructure and well-aligned community practices are essential foundations for reproducible research. Standards for identifying, describing, interlink and sharing digital objects are essential, but are almost 2,000 in number and often domain specific.[12] Achieving reproducibility at scale requires more than individual effort. It depends on shared standards, interoperable systems, and strong coordination across the research ecosystem. By investing in open, trustworthy infrastructures and fostering an active community committed to transparency and collaboration, stakeholders can create an environment in which research outputs are

[12] Source: FAIRsharing https://fairsharing.org/search?fairsharingRegistry=Standard

consistently findable, accessible, reusable, interoperable, and verifiable. TIER2 hence recommends that:

R1.1. Governments, institutions, and all types of funders should provide sustainable support for open infrastructures

Per the UNESCO Recommendation on Open Science (UNESCO, 2021), open infrastructures are an essential pillar of Open Science, and all stakeholders - including governments, funders and institutions - should ensure sustainable long-term support for the repositories, tools, standards, and identifier services that underpin reproducibility. Many critical components of the research ecosystem operate on fragile or short-term funding models, creating systemic risks for the preservation, accessibility, and interoperability of research outputs. As highlighted by the OECD Global Science Forum (OECD, 2025) and the Global Sustainability Coalition for Open Science Services (SCOSS[13]), if open infrastructures are to function as public goods, they require coordinated, reliable investment to remain trustworthy and community governed. Sustainable funding mechanisms and transparent governance structures are therefore crucial to guarantee continuity of services, enable community participation, and ensure that these infrastructures can evolve alongside research needs, ultimately strengthening the global capacity to verify, reproduce, and re-use scientific work.

R1.2. Research communities should improve and expand standards and guidelines for data re-use

While many existing frameworks focus on how researchers should produce reusable data, comparable guidance for those who *re-use* data remains less developed and is often fragmented across disciplines. Research communities, such as Academies, Councils, Associations and Societies, alongside communities focused around disciplinary infrastructures such as data and software repositories, should collaborate to develop and strengthen principles, standards and guidelines that support the responsible and effective re-use of data, with an emphasis on helping secondary users understand how to assess data quality, fitness-for-purpose, provenance, consent conditions, and methodological limitations. Although detailed standards should be domain specific, some overarching principles (e.g. ethical use, attribution, careful assessment of uncertainty, and concerns around data security and governance) may apply more broadly.

R1.3. Funders, publishers and institutions should require Globally Unique, Persistent and Resolvable Identifiers (GUPRIs) and open metadata for research objects and entities.

Global and openly resolvable persistent identifiers, alongside high quality and open metadata, are essential for reliably tracking research outputs—such as publications (including preprints), datasets, software, samples, pre-registrations, protocols and patents—and for connecting them to the people, organisations, projects, grants and other activities that produced them. Such connections form the backbone of a transparent, discoverable and interoperable research ecosystem, enabling others to find, retrieve, verify, re-use, and reproduce research results. To achieve this, funders and publishers of all types, as well as research institutions, should require, and implement where relevant, the use of Globally Unique, Persistent and Resolvable Identifiers

[13] https://scoss.org/

(GUPRIs)[14] across the research lifecycle, including (but not limited to) ORCID[15] for researchers, DOIs for datasets, publications and software, ROR IDs[16] for institutions, RAiD[17] for projects, and other community-endorsed identifiers. These identifiers must be accompanied by, and resolvable to, open, high-quality metadata that supports machine-readability, provenance tracking, and automated linking across systems. Mandating GUPRIs and open metadata strengthens research integrity, reduces ambiguity, supports FAIR data principles (Wilkinson et al., 2016), and enables consistent attribution and accountability across diverse research domains.

8.2. Incentives and policy

While many researchers recognise the value of transparent and rigorous practices, systemic pressures including publication expectations, career advancement criteria, and limited support and recognition for replication often act as disincentives. Effective policy must therefore align incentives with reproducible and trustworthy research, encourage communities to lead the development of appropriate norms, and ensure that assessment and monitoring systems reflect the diversity of research approaches across disciplines. Hence, to foster a culture in which practices that enable and deliver research reproducibility is not an additional burden but an integral and rewarded part of the research process, TIER2 recommends that:

R2.1. Funders, publishers and meta-researchers should develop and implement responsible metrics to enable monitoring of reproducibility practices in ways which account for the diversity of research approaches

Funders and other relevant stakeholders should strengthen monitoring of and communication about reproducibility to raise awareness among funded researchers, while recognising that any metrics used for this purpose must be carefully designed, responsibly implemented, and sensitive to the considerable epistemic diversity that exists across and within disciplines (Hicks et al., 2015). Reproducibility metrics can support automated oversight, for example by tracking availability of open outputs, but poorly designed indicators risk oversimplifying complex research practices, encouraging box ticking, disadvantaging fields with different norms, or becoming targets in their own right. Metrics might address both *enabling* (e.g. how transparent are research outputs) and actual *redoing* (e.g., the extent to which results from replication studies agree with initial findings) Metrics should complement, not replace, expert judgement; be transparent about their limitations; and be co-developed with and sensitive to the research communities they affect. By adopting reproducibility metrics that are field appropriate, proportionate, and used for learning rather than punitive evaluation, funders and other bodies can encourage continuous improvement in reproducibility practices without undermining the richness and diversity of research cultures.

R2.2. Funders should actively support and incentivise replication studies across all funding streams

Replication work remains undervalued in many fields, in part because standard grant mechanisms privilege novelty over verification. Yet replication is a core scientific activity that strengthens

[14] https://faircookbook.elixir-europe.org/content/recipes/infrastructure/gupri.html

[15] https://orcid.org

[16] https://ror.org/

[17] https://www.raid.org/

confidence in research findings and can guide improvements in methods, reporting, and research design across disciplines[18]. To address this gap, funders should mainstream replication by embedding explicit support across their portfolios of funding instruments. One approach is to embed replication in larger grants: proposals could begin by replicating the key prior studies their work builds upon. This grounds innovative research in verified findings while providing a principled basis for deciding which studies, among the many that could be replicated, most warrant the investment. Given that replication studies are still systematically undervalued, funders could also consider dedicated funding streams or top-up funding for well-justified replication studies. Such funding may stimulate early uptake, build community capacity, and signal the importance of verification as first-class research outputs. Over time, these efforts will support mainstreaming replication so it becomes a routine and expected part of funded research.

8.3. Training and skills

Improving reproducibility depends not only on strong policies and infrastructure but also on ensuring that researchers, supervisors, and editorial staff have the skills and support needed to implement good practices in their daily work. Building these capabilities requires coordinated investment in training, peer-support systems, and leadership development across the research ecosystem. By empowering individuals and communities, stakeholders can create a knowledgeable, confident workforce able to embed reproducibility as a routine and sustainable element of research practice. TIER2 hence recommends that:

R3.1. Institutions, funders, and publishers should build sustainable support networks and training ecosystems for reproducibility

Sustainable support networks and training ecosystems for reproducibility should be embedded within existing research and education structures, such as PhD programmes, postdoctoral training, and institutional professional development. Integrating coaching, peer-support mechanisms, and train-the-trainer models equips researchers at all career stages with practical, context-specific guidance and helps normalise reproducible methods early in their careers. Stakeholders should also invest in, and provide sustainable career pathways for, specialist roles (e.g., data stewards, statisticians) that support research teams with specific competences. In the long term, coordinated training ecosystems strengthen both individual and collective competences, as well as institutional capacity, ensuring that reproducibility becomes a routine and sustainable part of research practice.

R3.2. Institutions should strengthen leadership engagement and provide training to research leaders to foster reproducibility practices

Strengthening reproducibility requires active engagement from research leaders, supervisors, and principal investigators, who shape local research culture and set expectations for rigour, transparency, and documentation. Developing targeted training and incentive structures for these leaders can enhance their awareness of reproducibility challenges, equip them with practical strategies for supporting good practice within their teams, and reinforce their accountability as role models for responsible research behaviour. Embedding such training within institutional

[18] A leading example of such a funding scheme is that of the NWO Open Science NL Replication Studies Programme which will enter its third round in 2027. See: https://doi.org/10.5281/zenodo.17579089

frameworks—through leadership development programmes, supervisory training, promotion criteria, and departmental expectations—ensures that responsibility for reproducibility extends beyond early-career researchers and becomes a shared organisational priority. To avoid resistance from leadership, training programmes should be paired with awareness raising on why reproducibility practices matter when case training is obligatory, and with encouragement and incentives to take part, in particular when it is voluntary.

R3.3. Publishers should enhance journal capacity and infrastructure for checking and managing digital research objects

Publishers are facing growing challenges in ensuring research quality and provenance, including a documented surge in paper mill submissions (Richardson et al., 2025) and the widespread adoption of generative AI in scientific writing (Liang et al., 2024). To safeguard the integrity of published work, including its reproducibility, journals and publishers should strengthen their capacity to manage digital research objects by upskilling editorial and other relevant staff and dedicating resources to apply reproducibility and data-sharing checks consistently and effectively. These efforts must be supported by clear, regularly updated policies that reflect community standards and are openly registered in resources such as FAIRsharing[19]. As the volume and complexity of datasets, code, protocols, and other digital objects grows, editorial and other relevant teams (including research integrity, screening and some production teams) need budget, training, and technical support to assess compliance with data availability requirements, metadata standards, persistent identifier use, and repository best practices. To make these processes scalable and consistent, journals should collaborate with internal and external service providers to integrate automated and AI-enabled tools, such as automated data-availability checkers, code-execution verifiers, and metadata validators, directly into manuscript workflows. Given that the reliability of such tools is itself an active area of research, their viability should be regularly assessed. Where possible, cross-publisher collaboration on such services is to be encouraged. Ongoing initiatives, such as the STM Integrity Hub,[20] a shared, dynamic set of tools to screen manuscripts, as well as United2Act[21], a coalition of publishers to address the growing challenge of paper mills, offer very promising examples of such collaboration.

8.4. Strengthening the evidence-base

Effective policy for improving reproducibility must be grounded in robust empirical evidence. Yet despite growing recognition of reproducibility challenges, the evidence base for which interventions work, under what conditions, and with what trade-offs remains limited. A rapidly expanding meta-research community has begun to fill these gaps, but requires stronger support from funders, publishers, and institutions to generate actionable insights. TIER2 hence recommends that:

R4.1. Meta-researchers, supported by funders, publishers and institutions, should investigate the efficacy of reproducibility interventions

[19] https://fairsharing.org
[20] https://stm-assoc.org/what-we-do/strategic-areas/research-integrity/integrity-hub/
[21] https://stm-assoc.org/what-we-do/strategic-areas/research-integrity/united2act/

There remains a substantial lack of evidence on the outcomes of reproducibility interventions and on how interventions intersect, interact, or amplify each other. Most existing studies rely on proxy indicators aimed at *enabling redoing*, such as data-sharing rates or adherence to reporting guidelines, which offer only partial insight, as they rarely assess the success of *redoing* as direct outcomes (Dudda et al., 2025). Evidence on the effectiveness of training programmes is particularly limited, especially regarding which formats benefit which audiences and how long such effects persist. In addition, reproducibility is shaped by contextual factors that remain understudied, including epistemic and disciplinary norms, study populations, methodological traditions, and research team composition. To inform effective, evidence-based policy, meta-research is urgently needed to evaluate interventions rigorously, clarify how reproducibility manifests across diverse research communities, and identify the conditions that enable interventions to succeed. Funders should therefore provide targeted, sustained support for this work, recognising it as essential for improving research quality. In addition, publishers and funders should publicly share results of their internal investigations of reproducibility interventions.

R4.2. Meta-researchers, supported by funders and publishers, should investigate the costs and benefits of reproducibility interventions

Evidence on the actual costs and benefits of reproducibility interventions remains limited, making it difficult for policymakers to prioritise actions or allocate resources effectively. Future research should include systematic cost–benefit analyses that assess not only financial and time burdens for researchers, funders, and editorial/operational staff, but also impacts on research workflow, equity, and researcher wellbeing. It is equally important to examine the potential negative consequences of interventions, such as the risks of sharing low-quality or biased data, the possibility that standardisation may constrain methodological innovation, or the creation of new administrative burdens that disproportionately affect certain disciplines or career stages. By building a clearer evidence-base (using both qualitative *and* quantitative methods) on both the advantages and the trade-offs of reproducibility measures, meta-research can help funders, publishers, and institutions design effective policies and to implement workflows that are proportionate, targeted, and sensitive to disciplinary and contextual variation.

R4.3. Funders and publishers should enable meta-research regarding funding and publishing workflows through streamlined processes for collaboration and data-access

Meta-research on reproducibility frequently depends on access to funder and publisher data such as grant information, peer-review records, editorial decision timelines, and other publisher or funder workflows (e.g. screening checks), and often requires direct collaboration to run controlled trials of policy or workflow interventions. However, such research is currently hindered by fragmented processes, inconsistent data-sharing arrangements, privacy considerations, and unclear pathways or standards for establishing collaborations. Funders and publishers can therefore make a real difference by streamlining their internal workflows to support meta-research, as well as developing transparent policies on collaboration, standardised data-sharing agreement templates, and establishing secure mechanisms for researcher access to sensitive operational data. By making these processes more transparent and consistent, organisations can enable rigorous empirical investigation into how funding and publishing practices influence reproducibility and help generate evidence needed to inform effective policy-making.

Acknowledgements

We are grateful for the many stakeholders who took part in the many co-creative processes across our project. In particular, we thank colleagues who provided feedback throughout the co-creation of the TIER2 recommendations. We received invaluable input from colleagues within TIER2, our advisory board members, as well as our stakeholder communities across research funders and publishers. They helped in ensuring that our recommendations are relevant and feasible for the targeted stakeholders, clearly specified, and up to date with current development in the field.

References

Ahmed, H., & Lofstead, J. (2022). Managing Randomness to Enable Reproducible Machine Learning. *Proceedings of the 5th International Workshop on Practical Reproducible Evaluation of Computer Systems*, 15–20. https://doi.org/10.1145/3526062.3536353

Arroyo-Araujo M, Graf R, Maco M, van Dam E, Schenker E, Drinkenburg W, Koopmans B, de Boer S F, Cullum-Doyle M, Noldus Lpjj, Loos M, van Dommelen W, Spooren W, Biemans B, Buhl D L, & Kas M J. (2019). Reproducibility via coordinated standardization: A multi-center study in a Shank2 genetic rat model for Autism Spectrum Disorders. *Scientific Reports*, *9*, 11602. https://doi.org/10.1038/s41598-019-47981-0

Artrith, N., Butler, K. T., Coudert, F.-X., Han, S., Isayev, O., Jain, A., & Walsh, A. (2021). Best practices in machine learning for chemistry. *Nature Chemistry*, *13*(6), 505–508. https://doi.org/10.1038/s41557-021-00716-z

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

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

Bissell, M. (2013). Reproducibility: The risks of the replication drive. *Nature*, *503*(7476), 333–334. Scopus. https://doi.org/10.1038/503333a

Boettiger, C. (2015). An introduction to Docker for reproducible research. *SIGOPS Oper. Syst. Rev.*, *49*(1), 71–79. https://doi.org/10.1145/2723872.2723882

Buckley, J., Adams, L., Aribilola, I., Arshad, I., Azeem, M., Bracken, L., Breheny, C., Buckley, C., Chimello, I., Fagan, A., Fitzpatrick, D. P., Garza Herrera, D., Gomes, G. D., Grassick, S., Halligan, E., Hirway, A., Hyland, T., Imtiaz, M. B., Khan, M. B., … Zhang, L. (2022). An assessment of the transparency of contemporary technology education research employing interview-based methodologies. *International Journal of Technology and Design Education*, *32*(4), 1963–1982. https://doi.org/10.1007/s10798-021-09695-1

Büthe, T., & Jacobs, A. (2015). Conclusion: Research transparency for diverse discipline. *Qualitative & Multi-Method Research*, *13*(1), 52–64. https://doi.org/10.5281/zenodo.892983

Carter, S. M., & Little, M. (2007). Justifying Knowledge, Justifying Method, Taking Action: Epistemologies, Methodologies, and Methods in Qualitative Research. *Qualitative Health Research*, *17*(10), 1316–1328. https://doi.org/10.1177/1049732307306927

Chan, C., & Schoch, D. (2023). rang: Reconstructing reproducible R computational environments. *PLOS ONE*, *18*(6), e0286761. https://doi.org/10.1371/journal.pone.0286761

Chirigati, F., Rampin, R., Shasha, D., & Freire, J. (2016). ReproZip: Computational Reproducibility With Ease. *Proceedings of the 2016 International Conference on Management of Data*, 2085–2088. https://doi.org/10.1145/2882903.2899401

Cole, N. L., & Horbach, S. (2025). *TIER2 D1.4 Autoethnographic reflections on implementing radical reproducibility in the TIER2 project*. https://osf.io/nhak3

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* (No. n5zkw_v1). MetaArXiv. https://doi.org/10.31222/osf.io/n5zkw_v1

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

Dudda, L., Kormann, E., Kozula, M., DeVito, N. J., Klebel, T., Dewi, A. P. M., Spijker, R., Stegeman, I., Van den Eynden, V., Ross-Hellauer, T., & Leeflang, M. M. G. (2025). Open science interventions to improve reproducibility and replicability of research: A scoping review. *Royal Society Open Science*, *12*(4), 242057. https://doi.org/10.1098/rsos.242057

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

European Commission, Baker, L., Cristea, Ioana Alina, Errington, Timothy M, Jaśko, Katarzyna, Lusoli, Wainer, MacCallum, Catriona J, Parry, Vivienne, Pérignon, Christophe, Šimko, Tibor, Winchester, Catherine, & Lusoli, Wainer. (2020). *Reproducibility of scientific results in the EU: Scoping report*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/341654

Ferigo, D., Traversaro, S., Metta, G., & Pucci, D. (2020). Gym-Ignition: Reproducible Robotic Simulations for Reinforcement Learning. *2020 IEEE/SICE International Symposium on System Integration (SII)*, 885–890. https://doi.org/10.1109/SII46433.2020.9025951

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341). https://doi.org/10.1126/scitranslmed.aaf5027

Gundersen, O. E., & Kjensmo, S. (2018). State of the Art: Reproducibility in Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1), Article 1. https://doi.org/10.1609/aaai.v32i1.11503

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

Hamilton, D. G., Hong, K., Fraser, H., Rowhani-Farid, A., Fidler, F., & Page, M. J. (2023). Prevalence and predictors of data and code sharing in the medical and health sciences: Systematic review with meta-analysis of individual participant data. *BMJ*, e075767. https://doi.org/10.1136/bmj-2023-075767

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., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, *5*(8), 180448. https://doi.org/10.1098/rsos.180448

Hartley, M., & Olsson, T. S. G. (2020). dtoolAI: Reproducibility for Deep Learning. *Patterns*, *1*(5), 100073. https://doi.org/10.1016/j.patter.2020.100073

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Hopewell S, Ravaud P, Baron G, & Boutron I. (2012). Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: Interrupted time series analysis. *BMJ (Online)*, *345(7864) (no pagination)*, e4178. https://doi.org/10.1136/bmj.e4178

Horbach, S. P. J. M., Cole, N. L., Kopeinik, S., Leitner, B., Ross-Hellauer, T., & Tijdink, J. (2025). *How to get there from here? Barriers and enablers on the road towards reproducibility in research*. OSF. https://doi.org/10.31222/osf.io/gx9jq_v1

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, *4*(9), 100804. https://doi.org/10.1016/j.patter.2023.100804

Knorr Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press.
Kohrs, F., & Bannach-Brown, A. (2025). *TIER2 MS2.2 Self-reflections on co-creation activities*. https://osf.io/zc8uh
Laurinavichyute A, Yadav H, & Vasishth S. (2022). Share the code, not just the data: A case study of the reproducibility of articles published in the Journal of Memory and Language under the open data policy. *Journal of Memory and Language*, *125*, 104332. https://doi.org/10.1016/j.jml.2022.104332
Leitner, B., Tijdink, J., Kohrs, F., Bannach-Brown, A., Ulpts, S., Momeni, F., Adamidi, E., Klebel, T., Kormann, E., & Marangoni, A. (2025). *TIER2 D4.3 Pilot Implementation Report*. https://osf.io/7e6dy
Leonelli, S. (2018). Rethinking Reproducibility as a Criterion for Research Quality. In L. Fiorito, S. Scheall, & C. E. Suprinyak (Eds), *Research in the History of Economic Thought and Methodology* (Vol. 36, 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*, 1–21. https://doi.org/10.1017/psa.2022.45
Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., Cao, H., Liu, S., He, S., Huang, Z., Yang, D., Potts, C., Manning, C. D., & Zou, J. Y. (2024). *Mapping the Increasing Use of LLMs in Scientific Papers* (No. arXiv:2404.01268). arXiv. https://doi.org/10.48550/arXiv.2404.01268
Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
McDermott, M. B. A., Wang, S., Marinsek, N., Ranganath, R., Foschini, L., & Ghassemi, M. (2021). Reproducibility in machine learning for health research: Still a ways to go. *Science Translational Medicine*, *13*(586), eabb1655. https://doi.org/10.1126/scitranslmed.abb1655
Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596
Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M. H., Barbour, V., Coriat, A.-M., Foeger, N., & Dirnagl, U. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology*, *18*(7), e3000737. https://doi.org/10.1371/journal.pbio.3000737
Moreau, D., Wiebels, K., & Boettiger, C. (2023). Containers for computational reproducibility. *Nature Reviews Methods Primers*, *3*(1), 50. https://doi.org/10.1038/s43586-023-00236-9
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
OECD. (2025). *OECD Science, Technology and Innovation Outlook 2025: Policy Insights to Navigate Today's Shifting Landscape*. OECD Publishing. https://doi.org/10.1787/5fe57b90-en
Palmer, W., Okonya, O., Jellison, S., Horn, J., Harter, Z., Wilkett, M., & Vassar, M. (2021). Intervention reporting of clinical trials published in high-impact cardiology journals: Effect of the TIDieR checklist and guide. *BMJ Evidence-Based Medicine*, *26*(3), 91–97. https://doi.org/10.1136/bmjebm-2019-111309
Penders, B., de Rijcke, S., & Holbrook, J. B. (2020). Science's moral economy of repair: Replication and the circulation of reference. *Accountability in Research*, *27*(2), 107–113. https://doi.org/10.1080/08989621.2020.1720659
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

Pineau, J., Vincent-Lamarre, P., Sinha, K., Lariviere, V., Beygelzimer, A., d'Alche-Buc, F., Fox, E., & Larochelle, H. (2021). Improving Reproducibility in Machine Learning Research(A Report from the NeurIPS 2019 Reproducibility Program). *Journal of Machine Learning Research*, *22*(164), 1–20.

Pouchard, L., Lin, Y., & Van Dam, H. (2020). Replicating Machine Learning Experiments in Materials Science. In *Parallel Computing: Technology Trends* (pp. 743–755). IOS Press. https://doi.org/10.3233/APC200105

Prosser, A. M. B., Hamshaw, R. J. T., Meyer, J., Bagnall, R., Blackwood, L., Huysamen, M., Jordan, A., Vasileiou, K., & Walter, Z. (2022). When open data closes the door: A critical examination of the past, present and the potential future for open data guidelines in journals. *British Journal of Social Psychology*, bjso.12576. https://doi.org/10.1111/bjso.12576

Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1776–1826. https://doi.org/10.1145/3531146.3533231

Raste, S., Singh, R., Vaughan, J., & Nair, V. (2022). *Quantifying Inherent Randomness in Machine Learning Algorithms* (SSRN Scholarly Paper No. 4146989). Social Science Research Network. https://doi.org/10.2139/ssrn.4146989

Richardson, R. A. K., Hong, S. S., Byrne, J. A., Stoeger, T., & Amaral, L. A. N. (2025). The entities enabling scientific fraud at scale are large, resilient, and growing rapidly. *Proceedings of the National Academy of Sciences*, *122*(32), e2420092122. https://doi.org/10.1073/pnas.2420092122

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

Schlegel, M., & Sattler, K.-U. (2023). Management of Machine Learning Lifecycle Artifacts: A Survey. *SIGMOD Rec.*, *51*(4), 18–35. https://doi.org/10.1145/3582302.3582306

Schneider, J., Ulpts, S., Cole, N., & Ross-Hellauer, T. (2024). *TIER2 D2.5 - Policy Brief 1: Reproducibility and Epistemic Diversity*. https://osf.io/va45y

Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Haberl, A., Thalmann, S., & Kowald, D. (2025). Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Magazine*, *46*(2), e70002. https://doi.org/10.1002/aaai.70002

Sims M T, Bowers A M, Fernan J M, Dormire K D, Herrington J M, & Vassar M. (2018). Trial registration and adherence to reporting guidelines in cardiovascular journals. *Heart*, *104*(9), 753–759. https://doi.org/10.1136/heartjnl-2017-312165

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., Garijo, D., Grüning, B., La Rosa, M., Leo, S., Ó Carragáin, E., Portier, M., Trisovic, A., Community, R.-C., Groth, P., & Goble, C. (2022). Packaging research artefacts with RO-Crate. *Data Science*, *5*(2), 97–138. https://doi.org/10.3233/DS-210053

Stavropoulos, P., Lyris, I., Manola, N., Grypari, I., & Papageorgiou, H. (2023). Empowering Knowledge Discovery from Scientific Literature: A novel approach to Research Artifact Analysis. In L. Tan, D. Milajevs, G. Chauhan, J. Gwinnup, & E. Rippeth (Eds), *Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023)* (pp. 37–53). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.nlposs-1.5

Talkad Sukumar, P., & Metoyer, R. (2019). *Replication and Transparency of Qualitative Research from a Constructivist Perspective* (No. 6efvp_v1). OSF Preprints. https://doi.org/10.31219/osf.io/6efvp

Tijdink, J. K. (2023). *Future Studies—The Future(s) of Reproducibility in Research* [Pre-registration]. Open Science Framework. https://doi.org/10.17605/OSF.IO/WGTKU

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Tuval-Mashiach, R. (2021). Is replication relevant for qualitative research? *Qualitative Psychology*, *8*(3), 365–377. https://doi.org/10.1037/qup0000217

Ulpts, S., & Schneider, J. W. (2025a). *A conceptual review of uses and meanings of reproducibility and replication (Version 3)* (No. entu4_v3). MetaArXiv. https://doi.org/10.31222/osf.io/entu4_v3

Ulpts, S., & Schneider, J. W. (2025b). *Knowledge Production Modes: The Relevance and Feasibility of Reproducibility (Version 2)* (No. ujnd9_v2). MetaArXiv. https://doi.org/10.31222/osf.io/ujnd9_v2

UNESCO. (2021). *UNESCO Recommendation on Open Science* (p. 34). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en

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), 246. https://doi.org/10.1186/s13643-021-01780-9

Whitley, R. (2000). *The Intellectual and Social Organization of the Sciences* (Second Edition, Second Edition). Oxford University Press.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Xu, R., Baracaldo, N., & Joshi, J. (2021). *Privacy-Preserving Machine Learning: Methods, Challenges and Directions* (No. arXiv:2108.04417). arXiv. https://doi.org/10.48550/arXiv.2108.04417