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

Deliverable D2.5 -

Policy Brief 1: Reproducibility and Epistemic Diversity

31/12/2023

Lead Beneficiary: Know-Center GmbH

Author/s: Jesper Schneider, Sven Ulpts, Nicki Lisa Cole, Tony Ross-Hellauer

Reviewer: Thomas Klebel

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:	Policy Brief 1: Reproducibility and Epistemic Diversity
Deliverable n°:	D2.5
Version n°:	1.0
Nature of the deliverable:	Report
Dissemination level:	Public
WP responsible:	WP 2
Lead beneficiary:	Know-Center GmbH

TIER2 Project, Grant agreement No. 101094817

Due date of deliverable:	Month n° 12
Actual submission date:	Month n° 12

Deliverable status:

Version	Status	Date	Author(s)
0.9	Draft	15 Dec 2023	Jesper Schneider, Sven Ulpts, Nicki Lisa Cole, Tony Ross-Hellauer Aarhus University, Know-Center GmbH
0.91	Review	22 Dec 2023	Thomas Klebel Know/Center GmbH
0.92	Revised according to reviewer feedback	30 Dec 2023	JS, SU, TRH
1	Final (including final formatting and editing)	31 Dec 2023	JS, SU, TRH

EUROPEAN
POLICY BRIEF

Deliverable 2.5. Policy Brief 1

Reproducibility and Epistemic Diversity

Reproducibility is emerging as a key priority for research stakeholders. In this policy brief, targeted at research policy-makers (e.g., funders, institutions and publishers), we build on findings from the first year of the TIER2 project to deliver advice on creating policy that is sensitive to differences in basic conditions for knowledge production within and across fields (epistemic diversity) that crucially affect the relevance and feasibility of reproducibility.

31st December 2023

INTRODUCTION

Recently, concerns have arisen over purportedly low reproducibility rates across various disciplines, particularly in the behavioural and medical sciences. Proposed drivers of poor reproducibility include lack of transparency in reporting, data, and analysis, lack of replication studies, publication bias towards reporting of positive results, and questionable research practices. While low levels of reproducibility may pose a significant threat to scientific self-correction, the efficiency of research processes, and societal trust in research outcomes, the relevance of 'reproducibility' as an indicator of research quality differs greatly based on the nature of the research.

At first glance, the concept of 'reproducibility' seems intuitive, yet it conceals a labyrinth of complex, often contradictory interpretations. These variations in meaning significantly influence its application. Far from a unifying principle, 'reproducibility' is a multifaceted concept. The concept is embroiled in scholarly debate, reflecting a spectrum of understandings that can shift across disciplines, within them, or according to specific methodologies. Indeed, unique modes of knowledge production have developed their own norms and practices around 'reproducibility', if any, shaped by the nature of their research, methods and basic 'ways of knowing'. Its relevance and applicability are therefore not universal for all knowledge production modes.

Policies promoting reproducibility should therefore *not* perceive scientific activity in a monist way according to which there is a unique scientific method. Rather, such policies should respect that the basic conditions for knowledge production vary substantially within and across fields, which affects the development and/or understanding of knowledge.[1] The latter is known as 'epistemic diversity'. It is therefore important to recognise the particular challenges that each knowledge production mode faces, rather than mandate a universal approach to reproducibility. Ultimately, the adoption of reproducibility or transparency initiatives depends on social and political consensus within these modes.

TIER2 responds to these challenges by seeking to enable reproducibility where it is appropriate and caution against demands or incentives where it is inappropriate. We contribute to the following Horizon Europe Work Programme expected outcomes:

---

[1] Knorr Cetina, Karin. 1999. Epistemic Cultures: How the Sciences Make Knowledge. Cambridge, M.A.: Harvard University Press; Leonelli, Sabina. 2023. Philosophy of Open Science. Cambridgeshire, UK: Cambridge University Press.

- Expected Outcome 1: "Structured understanding of the underlying drivers, of concrete and effective interventions - funding, community-based, technical and policy - to enable 'reproducibility' of the results of R&I where it is appropriate; & of their benefits"
- Expected Outcome 2: "Effective solutions, policy-, technical- and practice-based, to enable the 'reproducibility' of R&I results where it is appropriate, in funding programmes, communities and dissemination"
- Expected Outcome 3: "Greater collaboration, alignment of practices and joint action by stakeholders to enable 'reproducibility' where it is appropriate, including but not limited to training, specialised careers and guidelines for best practice."

Hence, key to the project is to increase understanding of the (in)appropriateness of reproducibility as a quality criterion across different research contexts.

With this policy brief, we elaborate on this topic with reference to findings from the first year of the TIER2 project. Doing so is essential for structured understanding, effective solutions and successful joint action. We address those interested to formulate policy regarding reproducibility, including funders, institutions and publishers.

EVIDENCE AND ANALYSIS

With the mounting concerns about 'reproducibility' in the last decade, we have also seen a number of policy initiatives and guidelines that recommend mandating research practices assumed to promote reproducibility. Such initiatives are important, however, they are often presented, explicitly or implicitly, as one-size fit all solutions, where a specific notion of 'reproducibility' is promoted assuming it to be relevant and feasible for all kinds of research. Given that reproducibility, regardless of its definition, is only an appropriate quality criterion and practice depending on the nature of the research, promoting it as a universal indicator of research quality becomes problematic.

For example, the "Hong Kong Principles for Assessing Researchers"[2] recently influenced thinking about integrity and the design of codes of conduct in Europe. The principles incentivise specific practices linked to knowledge production in experimental fields like the behavioural and medical sciences. However, this approach is problematic for several reasons. First, the principles do not acknowledge the diverse range of research practices either within these fields or beyond. Second, they promote a general standard for assessing researchers, suggesting that researchers should be rated based on their adherence to certain tools designed to ensure a desired knowledge production. While the principles state that "diversity of types of research" should be valued, the interpretation of diversity is actually quite narrow. A certain notion of 'reproducibility' is evident throughout the principles, and the recommended tools support this view. Ultimately, this demonstrates the challenge of formulating principles or guidelines rooted in specific paradigms when seemingly the aim is to cover all types of research. If not, much more emphasis should have been given to addressing where and when such principles are relevant and feasible thereby acknowledging epistemic diversity.

A corresponding example is the scoping review on the "Reproducibility of Scientific Results in the EU"[3]. The review suggests that reproducibility should be seen as an ideal which policy makers should promote in order to increase quality, reliability and accurateness of research. The beliefs informing the review and its suggestions are again based upon a specific positivistic 'way of knowing', i.e., experimental research in the behavioural and medical areas. What is missing is a proper recognition of the potential differences in relevance of 'reproducibility' as an ideal across different 'ways of knowing', or the actual feasibility of implementing the numerous measures within areas where the notion of 'reproducibility' may be relevant. While disciplinary differences are acknowledged and suggestions for different interventions based on the presumed 'maturity' are suggested (p. 25), the basic premise is that all 'ways of knowing' essentially should have or strive for 'reproducibility' as an ideal, and this should be incentivised by funders and other stakeholders.

There are several other examples of general guidelines, principles or recommendations which all seem to propose top-down regulation of research practices based on particular 'ways of knowing' based on experimental methodology

---

[2] Moher, David et al. 2020. 'The Hong Kong Principles for Assessing Researchers: Fostering Research Integrity'. PLOS Biology 18 (7): e3000737. https://doi.org/10.1371/journal.pbio.3000737.

[3] European Union. 2020. 'Reproducibility of Scientific Results in the EU : Scoping Report.' Publications Office of the European Union. 1 December 2020. http://op.europa.eu/en/publication-detail/-/publication/6bc538ad-344f-11eb-b27b-01aa75ed71a1.

in the soft sciences. We suspect such regulation works best when the regulating entity (e.g., a journal, a professional organization, funder) shares the 'way of knowing' and the fitting knowledge production mode. Consequently, it is important that such guidelines or principles specifically address what they consider to be the target groups; and if they are supposed to be general, they should clearly acknowledge the apparent epistemic diversity among research practices and the consequences this has for the relevance and feasibility of reproducibility.

To foster increased awareness of epistemic diversity and its role in relation to reproducibility, the TIER 2 project has developed a framework that can help stakeholders to acknowledge these conceptual challenges.[4] Its aim is to help clarify specific types of 'reproducibility', their various functions, and relation to the specific nature of the research examined, e.g., a certain knowledge production mode, in order to examine the relevance and feasibility of reproducibility. Below we outline the framework.

Our extensive literature review on the meanings of reproducibility and related terms across the research landscape[5] suggests that:
1. The concept of 'reproducibility' 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. Hence, one size does not fit all.

'Reproducibility' is just one of many practices and criteria used to establish research quality (and accountability). However, contrary to what is typically conveyed in for example general guidelines or reproducibility initiatives, it is only suitable for certain types of research. Imposing a universal mandate or incentive for reproducibility by key stakeholders could therefore unfairly disadvantage certain types of research and the researchers involved. It may even lead to the exclusion of research, not because the research is of low quality, but because it is being held up to a quality criterion that is alien and potentially inappropriate in the first place. It is therefore essential to assess or at least consider the appropriateness of 'reproducibility' in relation to the research in question, before imposing requirements or incentives.

Principally, whether reproducibility is an appropriate quality establishing practice or criterion depends on the specific *type of reproducibility* in question and the *nature of the research* being conducted.

Type of 'reproducibility'
Due to the many faces of reproducibility and the substantial conceptual confusion, it has to be specified what is actually meant by 'reproducibility' for those examined, prior to evaluating its appropriateness. Types of reproducibility can be distinguished by two features:
1) **Practices:** What practices are involved, and what components of the research are addressed, which again can be distinguished between:
    a) *redoing* something and,
    b) *enabling* redoing (often in the form of transparency or sharing) or *enabling* general intersubjectivity (accountability).
2) **Functions:** What is the purpose with 'reproducibility'? What is supposed to be established, tested, or enabled?

Nature of the research
To understand how epistemic diversity affects the appropriateness of reproducibility, it is useful to differentiate between whether 'reproducibility' is epistemically *relevant* and practically *feasible*:

---

[4] Ulpts, Sven, and Jesper W. Schneider. 2023. 'Knowledge Production Modes: The Relevance and Feasibility of "Reproducibility"', MetArXiv. https://doi.org/10.31222/osf.io/ujnd9.

[5] Ulpts, Sven. 2023 Unravelling the conceptual and epistemic confusion surrounding the replications and reproducibility discourse. Open Science Framework. https://osf.io/4gsp6/

- **Relevance** considers whether redoing a study serves central functions like establishing causality, reducing errors, or enabling generalization. For example, reproducibility holds more relevance for clinical trials aiming to determine causal treatment effects than for an oral history study documenting personal experiences.
- **Feasibility** weighs practical constraints, accounting for needs around specialized expertise, data access, advanced equipment, and adequate funding. While relevant for gleaning insights about particle physics, reproducibility may simply be infeasible for many researchers given immense resource requirements.

Each of the two components of appropriateness can be assessed by looking at certain research characteristics:

- **Relevance:** Consider the way of knowing, the practices and criteria for establishing research quality (accountability), and whether commercial or proprietary goals conflict with redoing or enabling.
- **Feasibility:** Look at the nature of the subject of study, the uncertainty related to the research procedures, and the availability of the resources that the specific research and type of 'reproducibility' depend on.

Table 1. Research characteristics affecting the appropriateness of 'reproducibility'

Relevance	
Ways of Knowing	Research is always based on a certain way of knowing which determines what kind of knowledge claims are wanted, possible, and how they are produced. Different ways of knowing determine what quality establishing practices and criteria are relevant and how to apply them. Historically, 'reproducibility' originates in positivistic ways of knowing.
Quality establishing practices/ criteria	Depending on methodology and the methods used, the specific practices and criteria that are appropriate to establish the quality of the research vary.
Commercial & proprietary goals	Some research is not necessarily conducted by private companies but is driven by commercial goals or proprietary interests. These motivations can sometimes overshadow the objective of enabling others to understand or redo the research.
Feasibility	
Subject of Study	The subject of study can vary in its uniformity over time and environment. Therefore, it has to be considered how reactive the subject (phenomenon) is to its environment and to what degree it differs over time and location.
Uncertainty	The degree of understanding and agreement concerning how to apply procedures (methods, analysis, instruments, technologies) or investigate a subject differs between research communities.
Resource dependence and availability	Each investigation and practice require a specific investment of resources. The feasibility of conducting such an investigation or practice depends on how available such resources are for the people who are supposed to conduct it.

POLICY IMPLICATIONS AND RECOMMENDATIONS

Before formulating policies aimed at enabling or assessing reproducibility, stakeholders should ask themselves:

Who are the target group(s)?
- Consider what kinds of research should be included?
  - Be explicit about inclusion and exclusion

What do we mean with reproducibility?
- Specify what is meant by terms like reproducibility:
  - Differentiate between enabling and redoing
  - Specify what practices are involved (what is supposed to be done) and what is the intended function (what is intended to be tested, demonstrated, or enabled)

Is it appropriate for the research in question?
- Assess the relevance of reproducibility:
  - Look whether there are proprietary or commercial goals associated with the research that conflict with enabling others to know or even redo what was done.
  - It is crucial to identify what way of knowing the research is based on. Ask whether redoing or enabling makes sense for the way of knowing.

- See whether the specific type of redoing or enabling is an appropriate practice to establish quality and accountability for the type of research or whether there are already other practices fulfilling a similar or even more suited function.
- Assess the feasibility of reproducibility:
    - It is important to consider that the subject of study may not always be stable and uniform and can vary across different times and spaces. It also important to recognize that the degree of certainty researchers in a specific community have about the application of methods, procedures and instruments varies.
    - Finally, examine what resources the research depends on and how available those resources are for the specific type of redoing or enabling wanted.

Having reflected on the relevance and feasibility of reproducibility for the range of research covered by the publisher, funder, or institution, policymakers, two further initial considerations are especially pertinent for operationalising this in policy:

- Which elements of the intended policy can be broadly applied across all research (with exemptions or caveats), versus which elements of policy should be specifically targeted (e.g., to specific groups of journals or funding calls)?
    - It is essential to consider at which levels of granularity the policy should be designed. On the one hand, broadly applicable policies may be desired for reasons of simplicity and scale. For example, large publishers may wish to streamline the number of policies governing their portfolio of journals, or funders may wish to have requirements which are broadly applicable to all the research they fund, to aid understanding and simplify requirements amongst those whose research they publish or fund, or to streamline internal assessment practices. Yet overly broad policies will not do justice to the full range of research, for the reasons mentioned.
- Which elements of policy are already covered by existing Open Science or Research Integrity policies (e.g., data sharing, mitigation of questionable research practices)?
    - New policies should be designed to complement existing policies, especially to avoid potential conflicts between them.

Finally, our research reveals that the relevance and feasibility of reproducibility for qualitative research is heavily contested. While this is not a strict binary (there are areas of qualitative work that embrace certain aspects of reproducibility), nonetheless, based on our literature scoping and research activities, we recommend that:

- Policies targeting qualitative research should largely avoid the term 'reproducibility' (unless the term is explicitly endorsed by the epistemic community of researchers the policy targets, e.g., at the level of an individual journal). 'Reproducibility' is still a largely off-putting term to qualitative researchers because it points to an inappropriate expectation for their work. Hence, alternative terms like transparency, quality, rigour and credibility should be preferred.
- Qualitative traditions have their own long history of methodological transparency. Policies should seek to build upon and expand such traditions.
- Some specific practices require greater support and pilot-testing before broader roll-out. For example, sharing of qualitative data has well-known issues dues to the often highly sensitive nature of the data. There is hence a continued need for detailed guidance on how to prepare qualitative data for sharing, how to document the research process to make this data usable, and how to design consent as a living process that makes open data intentions and implications clear to participants. Pre-registration is also a practice which has been investigated as potentially enhancing transparency and trust in some qualitative research. Clearer guidance for qualitative researchers on not only how and why to use pre-registration, in which circumstances, is desirable.

SUSTAINABILITY AND LEGACY

TIER2 has planned an ambitious range of outputs, including toolkits, training, and guidelines, which will be announced during the project.

These resources will be hosted via *The Embassy of Good Science* to ensure their sustainability for the long-term.

See our project website for updates: https://tier2-project.eu/

TIER2 centres epistemic diversity by selecting three broad research areas - social, life, and computer sciences, and two cross-disciplinary stakeholder groups - research publishers and funders to systematically investigate reproducibility across contexts. Through coordinated co-creation with these communities, TIER2:

- Examines the epistemological, social, and technical factors that shape reproducibility across contexts.
- Builds a state-of-the-art evidence-base on the extent and efficacy of existing reproducibility interventions and practices.
- Co-creates techniques of scenario-planning, backcasting, and user-centred design to select, prioritise, adapt, and implement new tools to enhance reproducibility across contexts.
- Carries out a range of capacity-building activities, including fostering new Reproducibility Networks, awareness-raising campaigns, and training modules.

Through these actions, TIER2 contributes to increasing reproducibility where it is relevant and to preventing inappropriate demands for reproducibility where it is irrelevant or unfeasible.

PROJECT IDENTITY

PROJECT NAME	TIER2 (Enhancing Trust, Integrity and Efficiency in Research through next-level Reproducibility)
COORDINATOR	Tony Ross-Hellauer, Know-Center GmbH, Graz, Austria, tross@know-center.at
CONSORTIUM	Aarhus Universitet - AU - Aarhus, Denmark Athena Research & Innovation Center In Information Communication & Knowledge Technologies - ARC - Athens, Greece Biomedical Sciences Research Center Alexander Fleming - FLEMING - Athens, Greece The Chancellor Masters & Scholars of The University of Oxford (Associated Partner) - UOXF - Oxford, United Kingdom Charite, Universitätsmedizin Berlin - Charite - Berlin, Germany GESIS-Leibniz-Institut Für Sozialwissenschaften EV - GESIS - Mannheim, Germany Know-Center GmbH Research Center for Data-Driven Business & Big Data Analytics - KNOW - Graz, Austria Stichting VUmc - VUmc - Amsterdam, Netherlands Pensoft Publishing - PENSOFT - Sofia, Bulgaria OpenAIRE AMKE - OpenAIRE - Athens, Greece The Chancellor Masters & Scholars of The University of Oxford (Associated Partner) - UOXF - Oxford, United Kingdom
FUNDING SCHEME	HORIZON-WIDERA-2022-ERA-01-41 / Increasing the reproducibility of scientific results
DURATION	January 2023 – December 2025 (36 months)
BUDGET	EU contribution: 1 791 500 €
WEBSITE	https://tier2-project.eu/ https://twitter.com/TIER2Project
FOR MORE INFORMATION	Contact: Jesper Schneider, Aarhus University, jws@ps.au.dk Contact: Tony Ross-Hellauer, Know-Center GmbH, tross@know-center.at
FURTHER READING	TIER2 results are available from https://tier2-project.eu/library

- Ross-Hellauer, Tony et al. 2022. 'TIER2: Enhancing Trust, Integrity and Efficiency in Research through next-Level Reproducibility'. *Research Ideas and Outcomes*, 8. https://doi.org/10.3897/rio.8.e98457
- Ross-Hellauer, Tony et al. 2023. TIER2 Deliverable D3.1 – State-of-play on methods, tools, practices to increase reproducibility across diverse epistemic contexts. *Open Science Framework*. https://osf.io/cvsmw/
- Ross-Hellauer, Tony. 2023. 'Strategic Priorities for Reproducibility Reform'. *PLOS Biology* 21 (1): e3001943. https://doi.org/10.1371/journal.pbio.3001943
- Tijdink, Joeri et al. 2023. TIER2 Deliverable D4.1 – The Future(s) of Reproducibility in Research. *Open Science Framework.* https://osf.io/dzq9e/
- Ulpts, Sven, and Jesper W. Schneider. 2023. 'Knowledge Production Modes: The Relevance and Feasibility of "Reproducibility"', *MetArXiv*. https://doi.org/10.31222/osf.io/ujnd9.