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

Deliverable D1.4 – Autoethnographic reflections on implementing radical reproducibility in the TIER2 project

15/12/2025

Lead Beneficiary: Know Center Research GmbH

Authors: Nicki Lisa Cole, Serge P.J.M. Horbach
Reviewers: Sven Ulpts (Aarhus University), Eleni Adamidi (ARC), Tony Ross-Hellauer (KNOW)

Funded by the European Union

Sven Ulpts (Aarhus University), Eleni Adamidi (ARC), Tony Ross-Hellauer (KNOW)
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:	Autoethnographic reflections on implementing radical reproducibility in the TIER2 project
Deliverable n°:	D1.4
Version n°:	1.2
Nature of the deliverable:	Report
Dissemination level:	Public
WP responsible:	WP1
Lead beneficiary:	Know Center Research GmbH

TIER2 Project, Grant agreement No. 101094817

Due date of deliverable:	31/12/2025
Actual submission date:	19/12/2025

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft	28 November 2025	Nicki Lisa Cole (KNOW), Serge P.J.M. Horbach (Radboud)
1.1	Review	12 December 2025	Sven Ulpts (Aarhus University), Eleni Adamidi (ARC), Tony Ross-Hellauer (KNOW)
1.2	Final	15 December 2025	Nicki Lisa Cole (KNOW), Serge P.J.M. Horbach (Radboud)

Table of contents

Executive Summary

This report presents an autoethnographic study conducted within the TIER2 project, which committed to practicing and evaluating "radical reproducibility" and transparency in its own work. 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.

This 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 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.

List of Abbreviations

DMP – Data Management Plan
ECR – Early Career Researcher
EU – European Union
GA – General Assembly
OS – Open Science
OSF – Open Science Framework
FAIR – Findable, Accessible, Interoperable, Reusable

1. Introduction

In its research proposal (Ross-Hellauer et al., 2022), the TIER2 project promised that it would "adhere to radical reproducibility & transparency." The project aimed to be the change that it seeks by fostering a maximally open, transparent and reproducible approach to the implementation of Horizon Europe projects. This includes, among others, making its qualitative and quantitative data maximally findable, accessible, interoperable and reusable (FAIR), pre-registering and publishing full protocols for all scoping and pilot activities, using the containerised workflow tools that it creates to making data, code, software and algorithms open source and easily reproducible.

Implementation of such an ambitious agenda is not self-evident though. Especially in the context of a large consortium of partners spanning national, cultural, disciplinary, methodological and epistemological boundaries raises challenges related to diverse expectations, traditions and understandings of core reproducibility practices and concerns. Therefore, the project has committed to a reflective exercise in which its consortium members evaluate their own practices throughout the project, with the aim of better understanding and ultimately overcoming core challenges of reproducibility practices in the context of international and interdisciplinary research projects.

At the core of this reflective exercise is a project autoethnography of its own reproducibility practices. This task aimed to ensure best practices, including adherence to Horizon Europe requirements on Research Data Management and Open Science, and to take an autoethnographic approach to studying our own project as a case study. It served to reflect upon the challenges, costs and benefits resulting from such reproducibility and transparency approaches in the context of international, multidisciplinary project consortia.

2. Methodology

This study used an autoethnographic approach to studying reproducibility. Autoethnography is a qualitative research method in which researchers use autobiographical techniques to reflect on their own positionality, experiences and culture (Besio, 2009; Williamson, 2018). This can be done on an individual level or, as is the case in our study, on the collective or organisational level (Herrmann, 2020). This approach is particularly helpful in contexts of longitudinal self-reflection, where autobiographical techniques can be employed to trace development of meaning making and social and cultural embedding over time. In our case, these reflections mainly pertain to the development in understanding, challenges and benefits of reproducibility practices throughout the lifespan of a project that aimed to implement these. The study reported here consisted of three main components that generated the qualitative data analysed by the study:

1. Consortium-wide discussions at the project's kick-off meeting (in Jan 2023) and at each of its General Assemblies (GAs) (held in Jun 2024 and Jul 2025) to reflect upon challenges, expectations and experiences with regard to implementing reproducibility practices within the project.
2. Pre-GA surveys sent out to all consortium members to personally reflect on the practices to foster reproducibility of the project's work, as well as the potential challenges, costs and benefits involved (used to facilitate discussion at the kick-off meeting and first GA).

3. A series of quarterly diaries kept by a diverse set of five consortium members for the full duration of the project, reflecting on their reproducibility practices and perceptions, both related to TIER2 and beyond.

In total, these activities generated the following types and instances of data:

Table 1: Data types and instances

Data Type	Number of Instances
GA discussions (fieldnotes & recording)	3
Pre-GA surveys (text)	2
Reproducibility diary entries (text)	36

We will describe these components in more detail below.

2.1.GA discussions

During the project kick-off meeting and every subsequent GA, a 45-minute session was devoted to reflecting on reproducibility practices, potential and challenges within the TIER2 project. The session consisted of a plenary discussion moderated by the task leads of the autoethnography (the two authors of this report). These discussions took the answers to the pre-GA survey as input, except for the final GA, when a preliminary presentation of study results preceded the discussion session. Field notes were taken by the autoethnography task leads as a data source for the reflection report.

The discussions during these sessions were structured and consisted of multiple components, including discussion in pairs or trios, thinking about (one of) the tasks in TIER2 that they were or would be most involved with. The discussion questions for the kick-off meeting were anticipatory, and included the following:

- What reproducibility challenges do you anticipate? What challenges have you experienced?
- How did you address them or how could these be addressed?
- What would you need to reach maximal reproducibility of this task?

In addition, plenary discussions were held with the aim of collecting challenges, ideas and requirements for improved reproducibility practices. This included exercises like: Thinking about TIER2's core empirical tasks, who do we envision to reproduce these tasks or benefit from the efforts to optimise reproducibility (e.g. data or protocol availability)? What would these actors need to benefit maximally from our work?

In the second GA, discussions were of a more reflective nature, discussing how past practices aligned with TIER2 ambitions and identifying areas for potential improvement. Questions structuring the discussion included:

- What are your reproducibility highlights in TIER2 so far?
- What have been your reproducibility struggles?
- What are the discrepancies between what we have promised and what we are doing?
- What can we do better?

In the final project GA, a preliminary analysis of the autoethnography was presented to the consortium, inviting reflections on these findings, as well as allowing consortium partners to bring in additional reflections that they felt missing or misaligned with their personal experiences.

The following questions were used to guide a discussion session conducted in small groups and documented in a collaborative Google Doc:

1. How did TIER2 affect your understanding of reproducibility?
2. What went well, in terms of reproducibility, and should be recommended to future projects?
3. What tensions and/or conflicts around reproducibility have you experienced within the project?

Lastly, these sessions aimed to provide a platform for consortium members to voice any ideas or concerns they might have in relation to the transparency or reproducibility of the project's work. The presentations used to guide the discussions during the GA sessions are archived on Open Science Framework (OSF): https://osf.io/84wt5/overview.

2.2.Pre-GA Surveys

Ahead of each GA, consortium members were asked to fill out a short online survey reflecting on the reproducibility practices within the project. Data from these brief and individual reflections served as input into the plenary discussions during the GA. The survey was administered through the Padlet platform and contained the same questions as were outlined in the previous section for the kick-off meeting and the second GA. For the final GA, held in 2025, no pre-GA survey was administered, because the results of the preliminary analysis of the autoethnographic data were used as input to the GA discussions instead.

2.3.Reproducibility Diaries

Following the kick-off meeting, five consortium members volunteered to keep a 'reproducibility diary', reflecting on their reproducibility practices and perceptions, both related to TIER2 and beyond. The authors of these diaries were selected to achieve diversity of TIER2-tasks involved (i.e. they were involved in different work packages and tasks), academic seniority, and disciplinary background. They were asked to complete a diary entry every quarter, i.e. four times a year. Diary entries were free in form, length and content. They were collected through shared online documents that were only accessible by the diarist and the study team. In particular, diarists did not have access to each other's entries. Participants were given the following instructions:

"Please complete a diary entry reflecting on your ideals, challenges and practices related to reproducibility, both in your TIER2-related work as well as beyond. Please think of 'reproducibility' in its broadest sense; anything you feel to be related to it can be part of your diary entry. There are no restrictions regarding form, length or content of your contribution. Please just share whatever comes to mind. Topics you could write about include, but are not limited to:

- challenges you encountered in making your work reproducible or reproducing others' work;
- the value or meaning of reproducibility;
- discussions you had with others about reproducibility related issues;
- interesting ideas you heard or read that affected your perceptions about reproducibility, ways to achieve it and/or its desirability;
- practices you engaged in to increase reproducibility of your work; inspiring examples, best or worst practices you encountered related to reproducibility."

While we began with five diarists, not all were able to consistently participate. To maintain consistent data generation, the study leaders replaced one diarist mid-stream and were unable to replace another due to a lack of volunteers. Therefore, in total, we have data from six diarists, with four of them consistently making entries throughout the project duration.

A template for the diaries is included as part of our shared open materials on OSF: https://osf.io/84wt5/overview.

2.4. Data analysis

Our data analysis process was conducted in two phases. We first conducted a preliminary analysis of diary entries made until June 2025. These data were inductively and collaboratively coded by the authors to identify key themes within them. The preliminary results of this phase were fed into the final GA (see section 2.1). In the second phase, we coded the remaining diary entries and other data using the same coding scheme developed during the preliminary analysis. Due to the limited amount of data, we coded using comments in Google Docs rather than a coding software. The presentation of the preliminary analysis is shared as part of our open materials on OSF. In addition, during the second phase of coding we employed ChatGPT to deduplicate and restructure two long lists of synthesised data points for two categories of data: critical reflections and enablers of reproducibility. Understanding that ChatGPT can make errors and create fictional results, we cross checked its outputs with our original lists of data points as we reported the data, and are confident that the use of this AI tool made our process more efficient without compromising the integrity of our analysis and reporting. The ChatGPT requests and outputs are included in our supplementary materials, linked above.

3. Results

Here we present the results of our study, organised into five key themes: 1) learning outcomes, 2) positive reflections, 3) critical reflections, 4) tensions and conflicts, and 5) enablers of reproducibility.

3.1. Learning outcomes

Our findings regarding learning outcomes mainly come from diary entries but also include some reflections from the first and second GA. Consortium members discussed the meaning of reproducibility at the beginning and end of the project (first and third GAs), while diarists shared changes in their awareness of reproducibility and related topics and conceptual learning outcomes, shared "aha moments", and reflected on building skills through their work in TIER2.

We found that participating in the TIER2 project shaped consortium members' understanding of reproducibility. At the kick-off meeting, we observed that there was no consistent definition of reproducibility shared among the members, which is due in part to differing disciplinary and epistemic contexts. Yet, there was some consistency in members equating reproducibility practices with "good research" or with being a "good researcher". However, for some, in particular qualitative researchers, reproducibility was viewed as irrelevant to their work, and there was broader reflection on the need and usefulness of reproducibility across diverse epistemic contexts. This reflects the focus in TIER2 on centring epistemic diversity, and the aim of a task within the project, which was to map the conceptual framing of reproducibility across diverse disciplines (task 3.1, (see Ulpts & Schneider, 2023, 2024). Additionally, some members discussed reproducibility in terms of transparency (Open Science (OS) practices) and traceability, while others discussed it in relation to computing hardware.

Later, during the final GA, about six months prior to the project's conclusion, members reflected that they now understood reproducibility to be more complex than previously thought, both in terms of how it is defined and conceptualised, but also in terms of how it is operationalised. Reflecting the role of epistemic diversity in the role of the project, some members commented that they had a new understanding of the diversity and types of reproducibility (and replication) (this thread of conversation was also present during the second GA), and that they embraced and valued the complexity that our project illuminated. On the same theme, some shared an increased appreciation for considering the feasibility and practicality of pursuing "best practices" across diverse epistemic contexts, reflecting findings from the development of the conceptual framework of the project. Some noted that reproducibility is one aspect of doing good research but is not the only and it is not to be held up above all others, considering that it is not feasible for all. Related to this, some members shared an increased awareness of the differing value of reproducibility practices across diverse settings and career stages, and an increased appreciation for differing elements of openness and FAIRness, and the work burden that they generate.

Diarists shared that their understanding of OS practices had expanded through the project, with specific mention of their increased understanding of how to share data. They also developed greater awareness of reproducibility, what it is in a definitional sense, and what it means in practice. Reflecting both observations, diarists also learned the importance of creating a reproducible workflow or pipeline at the beginning of a research task, not only to improve its external transparency and reproducibility, but also to make updates to the data or analysis easier to implement, and to make adaption or deviation throughout the research process transparent. Diarists often reflected on their growth in awareness of the resources needed to implement reproducible and/or open research practices, and related to this, noted that it is necessary to weigh the costs and potential benefits of implementing (and improving) such practices. Finally, some diarists reflected that one can always do more or do better, in terms of making their research reproducible. Hence, being reproducible was considered not as some achievable end-goal, but as an ongoing process.

Members also shared practical learning outcomes. They reflected on learning through doing when implementing reproducibility practices within TIER2, developing reproducibility skills and believing that this made their work more rigorous. For example, some shared that they learned how to share data and code to facilitate peer review, how to better structure code for reproducibility, and how

to preregister a study protocol. Others remarked that they had developed the ability to identify research practices that undermine reproducibility, and that they were transferring their new skills to work outside of TIER2 by advocating for reproducibility practices in other projects. Still another reflected that they would implement reproducibility practices in the future based on what they learned within TIER2, and they believed that this would make their future research more efficient.

Finally, diarists shared some 'ah-ha moments' throughout the course of the project. Some diarists remarked that teams need consistent reproducibility training to collaborate and maintain a fixed standard, because inconsistent training leads to inconsistent practices within a team. Another shared that they came to view replication studies as a way to learn from others, remarking that conceptual replication "is a beautiful and creative way to study your topic." The same diarist suggested that the research world needs a "disruption" of existing norms and practices to achieve maximum research reproducibility.

3.2. Positive reflections

Consortium members expressed a variety of positive emotions and observations about the work of implementing reproducibility practices, studying them, and working to increase them within the TIER2 project. A few diarists expressed enthusiasm for the work, referring to it as "an exciting technical challenge" and expressing interest in the practical implementation aspects of it. Reflecting on highlights from the project during the 2025 GA, members observed good reproducibility practices had been implemented within the project, for example the consistent use of OSF throughout the project to share project workflows and outputs and generally working to be as open and transparent as possible. One diarist, remarking on the same theme, recounted efforts to implement "slow science" in their work, aiming to carefully document research workflows to make them transparent. Speaking to the work itself at the 2025 GA, members noted that the epistemically diverse community engagement and co-creative aspects of the project had worked well to develop strong tools and expressed value for the multistakeholder approach that was taken. Others noted that the process of evaluating the co-created tools had gone well, while others remarked on the positive collaborative spirit that members had cultivated throughout the project. In this way, reproducibility practices and the effort involved in exploring how to best implement them, served also social goals within the project consortium.

Members also expressed value for the work through the GA discussions of 2024 and 2025. Many remarked on valuing the reproducibility practices that they had implemented, which were recognised for structuring the work and improving the quality and value of the work by increasing transparency (e.g., preregistration of study protocols). Some shared that they had challenged themselves to "do better" and could see the results. Others shared that they found it rewarding to study reproducibility, and multiple diarists expressed gratitude for the reflection their entries stimulated.

Finally, pride was expressed by several members. Some felt proud of the work of the project, including the centring of epistemic and stakeholder diversity, and the perception that the project makes a valuable contribution to increasing reproducibility practices. Others expressed pride that our work is valued by those outside of the project.

3.3.Critical reflections, issues, tensions and conflicts

While learning outcomes and positive reflections were shared across the course of the project, the largest proportion of our data were critical in tone, occurring across all data types. In part, this is because our methods for the autoethnography task included prompts to reflect critically on the project and its progress. For example, during the kick-off meeting we asked “What reproducibility challenges do you anticipate? What challenges have you experienced?”. During the 2024 GA we asked, “What have been your reproducibility struggles? What are the discrepancies between what we have promised and what we are doing? What can we do better?”. During the final 2025 GA we asked, “What tensions and/or conflicts around reproducibility have you experienced within the project?” We therefore prompted more generation of data that are critical in nature than data which are positive (particularly during the 2024 GA discussion). Yet diarists, unprompted by us, also often shared critical reflections, and we observed tensions and conflicts around key discussion points during the GAs. In this section, we present these data organized around six sub-themes: 1) personal and emotional reflections; 2) time and resource constraints; 3) motivation and incentive structures; 4) methodological and technical challenges; 5) conceptual issues and epistemic tensions; and 6) consortium level and collaborative challenges.

3.3.1. Personal and emotional reflections

Across the data, we found that consortium members shared negative personal and emotional reflections on the challenges of implementing reproducible research, and our approach to studying it and attempting to increase it. Some shared anxiety about their own practices, questioning whether they are effectively “practicing what [they] preach”, and expressing hesitancy and a lack of confidence to share research workflows or outputs that may be perceived as “imperfect”. One diarist noted that “reproducibility is a great and dangerous concept, especially for perfectionists”. Some remarked that this could cause a leveling out in reproducibility practices, preventing some from advancing in this regard. Others expressed pessimism that others would attempt to replicate the work done within the project, and about the potential for our work to have impact beyond the “bubble” that we work within.

3.3.2. Time and resource constraints

Members expressed that the time and resources required to implement reproducible research posed a significant challenge. Time was flagged as an issue in terms of the time required make research outputs open, generally, and in particular for sharing qualitative data and making it FAIR, due to the consent process, de-identification, and documentation of the research context and process. Time-demand was also framed as an issue in terms of the competing interests that researchers have while working, like balancing multiple tasks or projects simultaneously, and doing coordination and management work. These factors led to members feeling that they have limited capacity to implement reproducibility practices, and some admitted that they were not consistently doing so. In light of this, the need to consider the costs and benefits of implementing reproducibility practices was a recurring conversation across the GA discussions, in particular regarding the qualitative research undertaken within the project.

Additionally, members noted that replicating research that used AI methods would require “enormous computational resources”, which was framed as a challenge to reproducibility. Relating to timing rather than time (as a resource), one member reflected during a GA that the timing of

writing the project's DMPs (M6, M18) was not ideal for reflecting current plans. A living document updated as plans evolved, this member suggested, could be more effective and valuable.

3.3.3. Motivation and incentive structures

Members reflected on issues with motivation, incentives, and engagement across the data landscape. Some remarked on the perceived lack of intrinsic motivation among researchers generally to implement reproducibility practices, and on a lack of extrinsic motivation within the research environment (e.g., incentives to implement Open Science practices), including a lack of recognition for reproducibility practices. In fact, we (the authors) observed that engagement with the self-reflection required for this study sometimes suffered from a lack of apparent motivation, indicated by inconsistent or absent diary entries and uneven engagement among members in the pre-GA surveys and discussion sessions. Some members reflected on disproportional engagement across generations, pointing out that there is a lack of implementation of reproducibility practices among senior research leaders and what some characterise as an "undue burden" placed on early career researchers (ECRs), who are often expected to implement them on behalf of teams, without training or mentorship. Some noted that there are systemic issues and entrenched behaviours that get in the way of implementing reproducibility practices.

3.3.4. Methodological and technical challenges

A key concern shared by consortium members across the data landscape is the potential transparency and traceability of the project's own work. Some shared their concern that we might provide insufficient technical detail for our methods to be reproduced, in particular that it would be challenging to provide sufficient documentation of the research process for qualitative tasks. Others noted the challenge that co-creative and conceptual aspects of the project pose for reproducibility, wondering how one could sufficiently document either type of process, or whether reproducibility should even be the goal for these. On the flip-side, one member observed that we may have blurred the lines in some tasks between planning for research and preregistering study protocols, questioning whether the time and resources invested for all preregistrations were necessary.

Members also noted that changes to methodological or technical aspects of the project's work could impact its potential reproducibility. For example, when research questions, methods and/or technical environments evolve over time. Some noted that a lack of technical knowledge gets in the way of implementing reproducibility practices. For example, one member remarked on the challenge of using a discipline-specific data repository without accompanying specialist knowledge.

These reflections testify to the effectiveness of the autoethnography task to stimulate consortium members to critically reflect on their own research practices in relation to reproducibility.

3.3.5. Collaborative challenges

Some members shared that collaboration posed as challenge in implementing reproducibility practices. Some noted that collaborating across disciplines or epistemologies led to some inconsistent implementation of reproducibility practices within task teams – also likely impacted by differing work styles. Others observed that this manifested as inconsistent practices of

documenting methods and analysis for some tasks and compromised our ability to open these materials for these tasks, thereby impacting the potential reproducibility of the work. Reflecting also various levels of knowledge of reproducibility practices, some members shared that they observed that some suboptimal practices had taken place within some teams, like failing to plan for data sharing for one task, for example.

Other collaborative challenges raised include the sense that we did not devote enough time to collaboratively contemplating, at the consortium level, the aims and objectives of the studies that we conducted. Some remarked that aligning the project ambitions, including its aim of practicing "radical reproducibility" or developing a "community of practice" were not always feasible, while others observed that some of the norms and values that we agreed early on were not uniformly maintained throughout the project.

Looking outward, some members stated that they felt we had insufficiently engaged with reproducibility and Open Science communities and similar projects that ran concurrently (with which there was some strategic collaboration). Yet, at the same time, some members shared at different points during the project that they feared that we were failing to reach outside of the "Open Science bubble", thereby limiting broader engagement of the research community with our project outcomes. Finally, near the end of the project, some shared concerns about the potential sustainability of project outputs (e.g., tools created through pilot activities) and systems.

3.3.6. Conceptual issues and epistemic tensions

A key conceptual issue mentioned in various reflection efforts concerns attempts to define reproducibility. When asked during the kick-off meeting what the term means to them, we found that there was no consistent definition offered by consortium members. Some associated it with transparency/traceability and/or Open Science practices; some associated it with hardware. Some claimed that they felt that reproducibility lacks relevance in the context of their work (e.g., qualitative researchers). We noticed a lack of conceptual clarity in use of terms like openness, transparency, and reproducibility. The absence of a universally shared definition at the start of the project foretold some of the project findings that would emerge through literature reviews (Cole et al., 2024; Dudda et al., 2023; Semmelrock et al., 2023; Ulpts & Schneider, 2024) and development of the conceptual framework for the project (Ulpts & Schneider, 2023) – there is in fact an absence of a consistent definition within the literature, across fields and disciplines, and epistemic contexts. The absence of agreement reflects the diversity of the project consortium.

Part of the challenge members experienced in attempting to understand, define and operationalise reproducibility is that there are key differences between what it means in quantitative versus qualitative epistemologies. When members stated that it was not applicable to their work as qualitative researchers, they noted that while transparency and rigour are considered quality criteria for their work, reproducibility is not, due to the often socially constructed, context-dependent, and researcher-laden nature of qualitative data and analysis. This view aligns with what was found by members who conducted a literature review focused on the definitions and concepts of reproducibility across diverse academic disciplines, and contributed to the development of their conceptual framework which posits that both the relevance and feasibility of reproducibility must always be taken into account before it can be deemed an appropriate quality

criterion or expectation for research. The discussion at the 2024 GA focused on this topic quite a bit, with one member remarking that when reproducibility is deemed irrelevant or infeasible, the reasons why this is the case should be given. This sparked the idea of a “reproducibility statement”, inspired by positionality statements in qualitative research that could be included when reporting on research.

Tensions emerged around the complicated nature of reproducibility during this same discussion. While some remarked positively that our collective work was revealing the complexity of reproducibility, some others felt that we were “overcomplicating things” by focusing on this complexity. These members expressed a desire to stick to simpler conceptualisations to move the work of increasing reproducibility forward. Still others suggested that the depth of discussion around this topic was a sign of how complex it in fact is, regardless of whether we want it to be. Particularly as regards epistemic diversity and applicability of reproducibility practices, one stated that it is important to discuss its complexity to better understand these dynamics.

This discussion signals tension around the operationalization of epistemic diversity within the project’s work. While some shared positive reflections and pride for the role that this concept and its operationalisation played within the project, others expressed frustration that it was not realized or operationalised as fully as it could have been. For example, while the project centered diversity by including life science, computer sciences and social sciences as fields of focus in tool development, this work encompassed only quantitative epistemologies.

Still, a few others indicated frustration with the embrace of epistemically diverse reproducibility and open practices. During both the 2024 and 2025 GA, some questioned whether it was appropriate or worthwhile that some reproducibility (i.e. Open Science) practices had been operationalised within qualitative aspects of the project’s research. While those who implemented these remarked positively that they had learned how to preregister qualitative studies, and had developed their knowledge and skills in opening qualitative data, methods and analysis, a few members stated that these practices were perhaps not worthy of the time and resources required to implement them, questioned whether they enabled transparency or reproducibility, and suggested that they may have been “performative” rather than effective in nature.

3.4.Enablers

Across the data landscape consortium members articulated what they thought could enable reproducibility practices, often in response to the critical reflections and challenges brought forward. Role models, mentoring and best practice exemplars were noted by some members as important enablers. Emphasised here was the importance of more senior researchers leading by example, to model best practices for junior researchers (instead of simply passing the work of implementing reproducibility on to them), though some also pointed at the potential educational value of *bad* examples. Members also noted the importance of communication, coordination and collaboration in implementing best reproducibility practices within a consortium. In terms of increasing reproducibility outside of the project, they emphasised the importance of translating research outcomes to external stakeholders. Reflections also touched on the strategy of implementing and increasing reproducibility practices, with some advocating for “sharing everything” while others countered that one should be pragmatic in deciding what to document,

considering whether an output or process requires heavy versus light documentation, and aiming only to implement practices required to meeting publishing and (data management plan) DMP requirement. Some noted the enabling role played by open-source software, and others advocated for a “slow science” approach of doing less but to a higher standard, to create the time and space for thorough implementation of reproducibility practices. The idea to implement “reproducibility statements” mentioned earlier was viewed as a way to increase engagement with reproducibility as a concept, especially among those who would otherwise eschew it as irrelevant to their epistemological approach. Members noted the enabling role of formal training, incorporated into education systems, and more attention to formal methods as another enabler. For example, members often mentioned the enabling role of documentation of the research process, and the role of technical aspects like containerisation of computational workflows and improved ways of de-identifying qualitative data. Members additionally discussed the need for “culture change” and the role of incentives to shift behaviour and produce this. Finally, members reflected on the need for better evidence of the efficacy of reproducibility practices (e.g., studies that demonstrate causation and accurate metrics) and of the cost-benefit of implementing them. Relatedly, members stated the need for better estimates of the costs and resources required for reproducibility practices, so that they can be properly funded. And members suggested that the TIER2 project should demonstrate to the broader community the efficacy of the tools it created, so that they will be taken up and assist in increasing research reproducibility.

4. Discussion and conclusion

This autoethnographic study of the TIER2 project found that studying reproducibility and aiming to implement it maximally was a positive and rewarding experience for the project consortium. Our findings show that both studying reproducibility and having a project commitment to “radical reproducibility” generated valuable learning outcomes, including the development of skills to implement reproducibility practices and the confidence to use them. We also found that this approach generated a multiplier effect, with the learning outcomes and skills developed by members being applied in other projects and funding proposals. Yet, some members experienced anxiety about revealing their work and potential mistakes through reproducibility practices, some experienced frustration due to epistemic conflicts with established reproducibility practices, and some felt pessimistic about the potential value of the project’s contributions. Despite these experiences, our findings suggest that the TIER2 approach was worthwhile as a knowledge, skill building, and amplifying exercise.

Our findings also illustrate the various factors that can get in the way of best reproducibility intentions. Though the consortium was composed of diverse experts on Open Science practices with experience in implementing them, we still encountered moments of uncertainty of how to proceed with reproducibility practices. Questions arose about how far to go or when to know when to stop, in a pragmatic sense, given the time and resources available, and some struggled with the collaborative implementation of reproducibility practices in multi-disciplinary teams. Though TIER2 was funded to study and increase reproducibility, our findings show that our consortium still suffered from a lack of time and resources to implement reproducibility practices with the consistency and efficacy that they wished to. One can reason that the lack of resources is a key factor in placing the “undue burden” on ECRs, who from a financial standpoint are “cheaper” than their more senior colleagues, which makes shifting reproducibility practices to them an

economically logical practice. These conclusions signal both the need for increased funding and training for reproducibility practices. Many members noted the importance of mentoring and example setting by senior researchers as important enablers of reproducibility. Members remarked that training in reproducibility practices should be a part of formal research training, but also, that senior researchers should be expected to engage in these trainings. Yet, knowledge of how to implement reproducibility practices will not necessarily offset their shift to junior staff if time is scarce among seniors. To open up time, it is perhaps necessary to consider the shift to "slow science", in which researchers intentionally take on less work in order to slow down their workflows and allow the time and space for reproducibility to flourish.

We see value in the complexity of reproducibility that was illuminated through the TIER2 project. While this complexity may have caused consternation among some at time, the work that highlights this shines an important light on considerations of epistemic diversity within discussions and innovations focused on reproducibility and Open Science. To that end, the creation of the conceptual framework of the project, which brings our attention to practices of "redoing" and of "enabling" redoing, helps to simplify the way we think about and approach the complex nature of reproducibility. It also makes it possible to incorporate diverse epistemologies into the conversation and workflows, by pushing us to consider the relevance and feasibility of reproducibility depending on epistemic context. This conceptualisation can lift the undue burden placed on scholars from the arts and humanities and qualitative researchers to abide reproducibility norms and practices that were designed for quantitative epistemologies. At the same time, our findings show that qualitative researchers within TIER2 were eager to implement Open Science and reproducibility practices in their project work (when relevant and feasible) and contributed to the expansion of the range of practices that were used within TIER2 as we sought to achieve "radical reproducibility". On that note, while limited time and resources are a known factor that must be accounted for when making pragmatic decisions about which reproducibility practices to implement when, the cost-conscious argument used by some consortium members to argue against some of the open qualitative research practices that were implemented during the project could serve to exclude non-quantitative researchers from these practices, from contributing to innovation in science reform, and from the ability to access the role of "good researcher" when that is equated with openness and reproducibility.

Finally, we note that the autoethnography task was effective in stimulating discussions and reflections about reproducibility practices and their implementation in an international, interdisciplinary research project. The collective discussions during the GAs provided a platform for collective reflection and required consortium members to articulate their understandings and perceptions regarding reproducibility or otherwise 'good' research practices. In addition, the personal reflections of the diarists were by several considered as helpful and rewarding exercises, allowing them to engage in reflections for which there is usually little time.

To conclude, we note that our study findings highlighted many of the same enablers of reproducibility that are already prominently accounted for in the existing literature base and science reform recommendations. Our findings add to this collective voice in calling for increased funding, resourcing, incentives and rewards to support the implementation of reproducible research. We also highlight the call by our consortium members for increased evidence of efficacy

of reproducibility interventions and practices, and evidence of their costs and benefits. The development of informed science policy and innovation cannot be undertaken without them.

Acknowledgements

The authors gratefully acknowledge the consortium diarists who donated their time to record their experiences and reflections across the three-year project duration and express their gratitude to all consortium members for their participation in annual reflection exercises. Their collaborative efforts strengthened the work of the project and created a rich set of reflections that we hope will be valuable to others who pursue research reproducibility in its various forms. We also gratefully acknowledged our partners who served as reviewers of an earlier version of this manuscript: Sven Arend Ulpts, Eleni Adamidi and Tony Ross-Hellauer. Their constructive feedback strengthened this report.

References

Besio, K. (2009). Autoethnography. In *International Encyclopedia of Human Geography* (pp. 240–243). Elsevier Science. https://www.sciencedirect.com:5037/science/chapter/referencework/abs/pii/B9780080449104004053

Cole, N. L., Ulpts, S., Bochynska, A., Kormann, E., Good, M., Leitner, B., & Ross-Hellauer, T. (2024). *Reproducibility and replicability of qualitative research: An integrative review of concepts, barriers and enablers*. OSF. https://doi.org/10.31222/osf.io/n5zkw

Dudda, L. A., Kozula, M., Ross-Hellauer, T., Kormann, E., Spijker, R., DeVito, N., Gopalakrishna, G., Van Den Eynden, V., Onghena, P., Naudet, F., Banzi, R., Fratelli, M., Varga, M., Gelsleichter, Y. A., Stegeman, I., & Leeflang, M. M. (2023). Scoping review and evidence mapping of interventions aimed at improving reproducible and replicable science: Protocol. *Open Research Europe*, *3*, 179. https://doi.org/10.12688/openreseurope.16567.1

Herrmann, A. (Ed.). (2020). *The Routledge International Handbook of Organizational Autoethnography*. Routledge. https://doi.org/10.4324/9780429056987

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

Semmelrock, H., Kopeinik, S., Theiler, D., Ross-Hellauer, T., & Kowald, D. (2023). *Reproducibility in Machine Learning-Driven Research* (No. arXiv:2307.10320). arXiv. https://doi.org/10.48550/arXiv.2307.10320

Ulpts, S., & Schneider, J. W. (2023). *Knowledge Production Modes: The Relevance and Feasibility of 'Reproducibility'*. MetaArXiv. https://doi.org/10.31222/osf.io/ujnd9

Ulpts, S., & Schneider, J. W. (2024). *A conceptual review of uses and meanings of reproducibility and replication*. https://doi.org/10.31222/osf.io/entu4

Williamson, K. (2018). Ethnographic research. In *Research Methods* (Second, pp. 311–335). Chandos Publishing. https://doi.org/10.1016/B978-0-08-102220-7.00013-3