Checklists for Computational Reproducibility in Social Sciences: Insights from Literature and Survey Evaluation

Fakhri Momeni*
Knowledge Technologies for the Social Sciences
GESIS - Leibniz Institute for the Social Sciences
Köln, NRW, Germany
fakhri.momeni@t-online.de

Muhammad Taimoor Khan*
Knowledge Technologies for the Social Sciences
GESIS - Leibniz Institute for the Social Sciences
Köln, NRW, Germany
taimoor.khan@gesis.org

Johannes Kiesel
Knowledge Technologies for the Social Sciences
GESIS - Leibniz Institute for the Social Sciences
Köln, NRW, Germany
Johannes.Kiesel@gesis.org

Tony Ross-Hellauer
Know Center Research GmbH
Graz, Austria
tross@know-center.at

Abstract

The use of AI-based computational models has increased significantly in social science. As in other disciplines, it is crucial for the reliability and transparency of the research results that they can be reproduced using publicly accessible resources. In social science, however, data access, ethical restrictions, non-standardized methods, and inadequate documentation procedures require additional measures to ensure reproducibility. In this paper, we address these challenges through three main contributions. First, to assess the current state of computational reproducibility across disciplines and identify gaps specific to social science, we conducted a systematic literature review of 42 peer-reviewed publications from the Scopus database. Second, to understand current practices and challenges in reproducibility, and to evaluate the checklist items identified from the literature, we conducted two surveys having 180 and 64 social science participants, respectively. Third, to address the reproducibility challenges of the social science community, we designed computational reproducibility checklists based on the aforementioned literature review and surveys. The checklist items received strong community support as 98.43% of the survey participants favored 76.35% of the checklist items for inclusion in the final checklists, particularly data documentation, source sharing, and methodological reporting received strong agreement. The checklists contain clear and actionable directives that are aligned with research processes to support efficient integration with reproducible social science workflows. The presented checklists are already employed to enhance the reproducibility of models on the *Methods Hub*[1] portal.

CCS Concepts

• **Theory of computation → Machine learning theory**; • **Human-centered computing → User studies**.

Keywords

computational reproducibility, social sciences, reproducibility checklist

ACM Reference Format:
Fakhri Momeni, Muhammad Taimoor Khan, Johannes Kiesel, and Tony Ross-Hellauer. 2025. Checklists for Computational Reproducibility in Social Sciences: Insights from Literature and Survey Evaluation. In *ACM Conference on Reproducibility and Replicability (ACM REP '25), July 29–31, 2025, Vancouver, BC, Canada.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3736731.3746161

1 Introduction

The contemporary research landscape increasingly emphasizes open research to ensure its quality. Open research includes the public sharing of research resources under open licenses, thereby facilitating the re-use of existing resources, establishing the trustworthiness of research, and fostering a culture of collaboration. The FAIR principles (findable, accessible, interoperable, and reusable) encourage researchers to adopt open science practices and ensure that their data, methodologies, and publications are accessible [24]. Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have catalyzed interdisciplinary collaborations across fields such as computer science with social sciences. These collaborations have yielded groundbreaking results by using state-of-the-art AI/ML approaches, which have become prominent drivers of innovation. However, the recent progress has also underscored significant reproducibility challenges in computer science [16] and the use of computational models in other disciplines [27]. Although open science and research reproducibility concepts are gaining momentum, many studies are not reproducible due to a lack of information or ambiguity in the information to reproduce them. Even the publications that claim to be reproducible are often found to be not fully reproducible [3]. Generally, sharing the research resources, access protocols, funding sources, and statement of interest adds to

*Both authors contributed equally to this research.
[1] https://methodshub.gesis.org/

ACM REP '25, Vancouver, BC, Canada

ACM ISBN 979-8-4007-1958-5/25/07
https://doi.org/10.1145/3736731.3746161

the research credibility [21, 42]. While external validity indicators like preregistration are showing signs of misuse [47].

Reproducibility is a complex concept that has varying interpretations in different disciplines. It is fundamentally defined as the ability of independent investigators to achieve the same outcome from an experiment by following the documentation provided by the original authors [20]. A more nuanced understanding of reproducibility is beyond binary, i.e., outcome reproducibility (reproducing the same results), analysis reproducibility (getting the same interpretations from different results), and interpretation reproducibility (drawing the same conclusions from different interpretations). The investigator may be the original authors, i.e., 1°CR (first-order computational reproducibility), an independent and trusted third party, e.g., a journal editor, i.e., 2°CR (second-order computational reproducibility), or the public at large, i.e., 3°CR (third-order computational reproducibility [53]. Furthermore, the resources may be publicly available to all, accessible through an authentication-based system, or explicitly requested. Additionally, as a tier system, reproducibility and its elevated levels, i.e., replicability, robustness, and generalizability, may not be targeted altogether [53]. An uncritical pursuit of reproducibility as a universal epistemic value can be misleading and therefore difficult to achieve [35]. [30] has also raised concerns over the level of reproducibility and their epistemic factors at each research stage, while also questioning the feasibility of its practices. Therefore, clear planning and adhering to reproducibility or its elevated tiers throughout the research process are essential. In this study, we limit our scope to analyzing computational reproducibility, i.e., concerning computational models only, aiming for 3°CR through publicly available resources in the social science discipline. The organizational and communal challenges and their counter incentives that are not specific to computational reproducibility and social sciences are not considered in this study [2, 34].

Computational reproducibility issues are reported across different disciplines. There are technical barriers, i.e., software configurations, package versioning, commands history, etc., and procedural barriers, i.e., lack of model sheets, checklists, and reporting practices [56]. In biomedicine, errors in code, differences in results, and insufficient documentation are common reproducibility barriers [51]. In computational communication science, only 5.88% of studies explicitly address the general ethical considerations [29]. Social science research is also affected by common obstacles such as dependency updates, software version changes, and the inability to perfectly restore the experiment environment [34]. Along with the discipline-agnostic computational reproducibility challenges, social science has specific barriers. The research data that is often collected from digital media platforms is dynamic and continuously evolving. For example, in a longitudinal assessment of 30 recollected Twitter datasets, 18.6% tweets and 20% users were found missing, when hydrated years later [64]. The privacy and ethical restrictions limit data sharing, e.g., the actual tweets in the dataset cannot be publicly shared and are instead hydrated from their unique IDs, when needed. The lack of open research tools, the prevalent use of proprietary and commercial software, e.g., SPSS and Stata, compatibility issues among software versions, and differences in computing resources also hinder computational reproducibility in social science [51]. Therefore, despite the growing emphasis on making research resources publicly available, concrete measures are needed to address the issues highlighted for improving computational reproducibility in social science [11].

While reproducibility is widely discussed in psychology and political science, it lacks tailored frameworks and checklists in social science [40]. Thus, building on existing efforts in related disciplines, we identify concrete checklist items from computational reproducibility literature. They focus on addressing the technical, procedural, and data access requirements in social science. The checklist items are refined, adapted, and prioritized through community feedback in two surveys on present practices and challenges, and the necessity of the identified checklist items in preserving computational reproducibility. The surveys are empirically evaluated, resulting in splitting the checklist items into three checklists based on their relevance to the research stage, i.e., data access, analysis, and sharing and archiving. This paper contributes:

- A systematic literature review of peer-reviewed publications from the Scopus database, categorizing computational reproducibility practices across disciplines and identifying gaps specific to computational reproducibility.
- A large-scale online survey of 180 participants to assess the current computational reproducibility practices and challenges faced in the social science community.
- A focused survey with 64 participants to validate and evaluate checklist items (59) identified from the literature.
- Refined, adapted, prioritized, and finally organized the checklist items into three checklists.

The checklists offer a comprehensive and actionable mechanism aligned with research stages to improve computational reproducibility in social science. In addition, the checklists are assisted by guides and reporting templates to improve transparency and checklist adoption by reducing human effort and errors.

2 Methodologies

In this study, we employ a two-method approach to identify and validate checklist items for computational reproducibility in social science, depicted as a methodological workflow in Figure 1. The first method involves a systematic literature review to identify checklist items from computational reproducibility gaps in social science and relevant efforts in other disciplines. It resulted in a list of 59 checklist items focusing on data access, technical preservation, and reporting standards. The second method involves two surveys conducted to gather community feedback on the state of computational reproducibility practices and challenges and to evaluate the identified computational reproducibility checklist items. Empirical analysis of the two surveys translates into the final checklists.

2.1 Systematic Literature Review

The primary resource for literature search is the 2023 version of the in-house Scopus database, maintained by the German Competence Centre for Bibliometrics (Scopus-KB). The search relied exclusively on this database to collect relevant articles. This choice was made due to Scopus' extensive multidisciplinary coverage, providing a robust foundation for identifying relevant literature on computational reproducibility. As computational research is fast

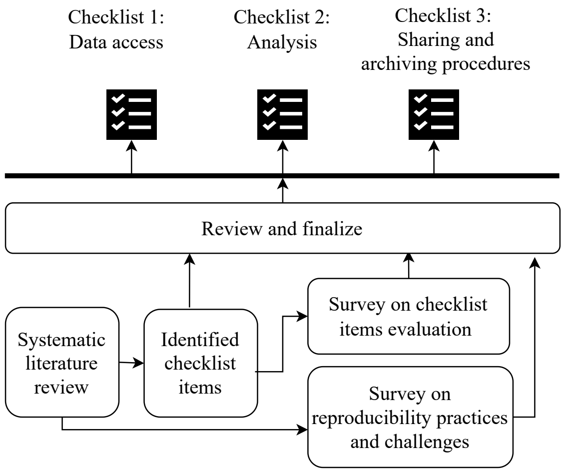


Figure 1: Methodological workflow of our two-method approach identifying computational reproducibility checklist items from literature, and improving them through surveys.

evolving, to focus on the most recent developments in computational reproducibility, we only searched for papers published in 2018 and onward. We followed a structured and systematic approach to collect relevant articles using search terms, i.e., *"computational reproducibility"*, *"computational reuse"*, *"computational replicability"*, *"replicable code"*, *"reproducible code"*, *"code reproducibility"*, and *"computational sustainability"* in the title and keywords of the article. Articles containing one or more search terms in their title or keywords are considered across disciplines to maximize the breadth of the collection. Aligning the collection with our research focus, the articles that do not discuss, evaluate, or enhance computational reproducibility are filtered. Further, duplicate articles were also filtered from the final collection.

In the initial review of the articles, we extracted their metadata, key themes such as research discipline, presence of intervention or evaluation, reproducibility outcomes, and methodological approach. Based on this, we then categorized the articles into four types: (1) reproducibility experiments, (2) reproducibility intervention, (3) post-intervention review, and (4) comparative analysis. The foundational understanding of computational reproducibility literature served two purposes. First, it led to the formulation of key questions about the current practices and challenges observed in the literature. Second, it helped in reviewing checklist items across disciplines. Both aspects are incorporated into surveys to gather community feedback.

2.2 Survey Design

Participants recruitment:

The target group for the two surveys is social scientists working on computational approaches. In order to identify this target group as surveys participants, we explored the Scopus database for the corresponding authors of recent publications (i.e., 2016–2023) containing relevant keywords[2]. We used the EUSurvey platform[3] to design the surveys, gather responses, and evaluate them. The participants are invited to both surveys simultaneously (in July 2024) and may respond to one or both surveys at their discretion. The invitation acknowledges the effort required to respond to the second survey that involves rating the checklist items, and therefore, additional time was allocated to complete it. The surveys were reviewed for ethical and data protection compliance by the legal affairs and data protection office at *GESIS - Leibniz Institute for the Social Sciences* for not gathering sensitive private information and rather solely relying on publicly available information. The participants contributed voluntarily as anonymous respondents and were explicitly informed of the intended purpose of the study and the type of information collected.

The invitations for both surveys were disseminated to approximately 26,000 potential respondents from the target group, i.e., social scientists with a publication track on computational models. The first survey on reproducibility practices and challenges received 180 responses with a response rate of 0.7%. The second survey on checklist items evaluation received 64 responses with a response rate of 0.25%. Although the survey participants represent a small sample in the social science community, however, they are carefully selected based on their direct involvement in the topic. It suggests that the responses are informed and contextually valid. Moreover, the surveys were intended to gather exploratory insights and were not necessarily designed to be statistically representative of the entire population.

Survey on reproducibility practices and challenges: The first survey aimed to assess current practices, attitudes, and challenges related to computational reproducibility in social science. It included multiple-choice and Likert-scale questions covering open science practices, documentation strategies, perceived obstacles, and desired features for reproducibility support tools. This survey received responses from 180 participants, representing a diverse set of roles, including faculty members, postdoctoral researchers, and PhD students. However, senior academics with over a decade of experience and serving in higher designations represented the dominant group among the participants.

Survey on checklist items evaluation: The second survey focused on gathering responses on 59 checklist items identified from the literature review and existing resources. It used a 9-point Likert scale rating (with 1 as strong disagreement and 9 as strong agreement) to evaluate the necessity of the checklist items for computational reproducibility in social science. The checklist items are designed to be applicable across common social science workflows using open-science tools and public data. This survey does not include open questions. This survey received responses from 64 participants, having highly overlapping demographics with the participants of the other survey. However, due to the anonymity of responses, it cannot be confirmed whether the same individuals

[2]The keywords used for author identification are: *"computational social science"*, *"digital sociology"*, *"social data science"*, *"computational sociology"*, *"social network analysis"*, *"big data social science"*, *"data-driven social science"*, *"computational behavioral science"*, *"quantitative social science"*, *"social informatics"*, *"digital humanities"*, *"computational anthropology"*, *"socioinformatics"*, *"e-social science"*, *"computational politics"*, *"algorithmic social science"*, *"social simulation"*, *"agent-based modeling in social science"*, *"social media analytics"*, *"computational economics"*, and *"computational demography"*.
[3]https://ec.europa.eu/eusurvey

participated in both survey rounds. Based on the feedback from the two surveys, the final checklist items have been decided and organized into three checklists.

3 Systematic Literature Review on Computational Reproducibility

To ground our work in existing research, we conducted a systematic literature review of publications on computational reproducibility across disciplines. The goal was to analyze existing contributions on computational reproducibility in related disciplines as well as identify the common practices and tools, and the associated barriers and challenges in social science. Following the literature review design in 2.1, we initially collected 75 articles that were reduced to 42 articles after the filtering process. We then extracted the information relevant to reproducibility from these articles, including research objectives, the computational method used, intervention type, contributing tools, guides or frameworks, conducting reproducibility experiments, and discussing disciplinary context-based findings. Based on research focus and type of contribution, these articles are organized into 4 categories, i.e., reproducibility experiment, reproducibility intervention, post-intervention review, and comparative analysis. Table 1 provides representative articles in each of these categories. The categorization and category naming are influenced by the terms and explanations used in the articles.

- *Reproducibility experiment:* It includes reproducibility studies on selected papers that claim to be reproducible. The original results are being reproduced using publicly available resources and documentation.
- *Reproducibility Intervention:* These studies introduced interventions to tackle open challenges in computational reproducibility. The interventions may be tools, process guides, or reporting standards.
- *Post-Intervention Review:* These studies conduct demonstrations to analyze the effectiveness and utility of the available interventions. The impact of the interventions available is quantified by comparing the reproducibility metrics before and after the intervention was introduced.
- *Comparative Analysis:* These studies provide a comparative analysis of the reproducibility interventions, i.e., tools, software, frameworks, or conceptual approaches, along with different direct and indirect measures of reproducibility.

3.1 Reproducibility Experiment

The contributions to reproducibility experiments mainly focus on sharing research resources and reproducing original results using available resources. A lack of adherence to open science practices is observed in plant pathology [59]. Psychology has a relatively higher rate of data and code availability; however, it lacks an overall strategy to utilize these resources [44]. [3] reported challenges with data retrieval and author responsiveness in wildlife sciences, indicating challenges in data sharing practices. [54] pointed out difficulties in reproducing longitudinal analyses due to the lack of open science practices and guides to demonstrate their use. In restoration ecology, inconsistencies in data and methodologies are demonstrated [8]. In some cases, code obsolescence issues are identified where a study showed 7 out of 20 contributions from the Institution for Social and Policy Studies (ISPS) data archives have code errors [46]. In general, there is a growing consensus across disciplines that the sharing of data and code is essential but does not guarantee computational reproducibility [60].

3.2 Reproducibility Intervention

Several studies have proposed interventions to address the reproducibility shortcomings identified in reproducibility experiments. They mainly consist of practical tools, guides, and reporting standards to demonstrate improvement in computational reproducibility.

Tools: Tools assist researchers in making their contribution reproducible as an aided functionality, automating manual procedures, and reporting for consistency. They minimize human involvement, reducing both effort and manual errors. Tools may keep track of resource provenance and data performance [48], encapsulate data and code [10], classify datasets and software from academic documents to their source URLs [50], and integrate the Dataverse platform to reproducibility platforms [61]. Docker containers and cloud computing [36], integrating replication-package metadata, functionality APIs, and virtual containers [61] are advised. The role of virtual containers is emphasized beyond reproducibility for collaboration, scalability, and research management while acknowledging its potential limitations in high-performance computing [41]. Time-based code with data capsules may be preserved for computational reproducibility, promoting transparency with minimal integration overhead [45]. To overcome technical barriers, *EasyScienceGateway* provides means for non-technical users to manage their research environment with a better user experience [38].

There are reproducibility tools and frameworks designed for specific disciplines, e.g., to quantify numerical uncertainty in neuroimaging [49], research data management for microbial communities [17] and semantic mediation in biomolecular data [19]. In hydrology, reproducibility tools facilitate the sharing, verification, and publishing of hydrologic data and models [15, 58]. Jupyter is scaled for the analysis of geospatial data [62]. [7] and [25] explored containerization and cloud computing to achieve reproducibility in Geospace and ocean modeling. [5] and [58] addressed the computational demands of large datasets and sustainability issues in prediction systems. In recent research, there is growing recognition of attaining computational reproducibility and the technical complexity associated with it. Computational research in social science needs specialized tools and flexible frameworks to use these tools, which can also adapt to the dynamics of research resources and demands.

Process Frameworks: Reproducible processes and workflows for computational methods emphasize detailed documentation and their open access [14]. The computational reproducibility pitfalls are identified and addressed by adapting version control and dependency management tools for practical solutions in the signal processing domain [57]. The need for practical guides to control the use of reproducibility tools is examined in spatial statistic surveys [31]. [33] reviewed advances in functional neuroimaging, highlighting both the progress made in statistical analysis and the challenges

Table 1: Categorization of literature on computational reproducibility.

Category	Representative Articles
Reproducibility Experiment	[59], [46]
	[60], [44], [3], [55], [8]
Reproducibility Intervention	[62], [49], [5], [48], [19], [36], [15], [45], [58], [7],
	[38], [25], [10], [17], [61], [41], [50]
	[39], [57], [14], [4], [31], [33] [23], [43]
Post-Intervention Review	[12], [13], [37], [54]
Comparative Analysis	[32], [28], [18], [1], [26], [22]

posed by big data. Teaching computational reproducibility is advised for educational benefits, offering practical strategies as a conceptual framework for instructors to integrate reproducibility into their curricula [4]. A project-based computational reproducibility course is also suggested, focusing on specific challenges of neuroimaging and the available tools to address them. Collectively, these studies underscore that the effort required to achieve computational reproducibility can be significantly reduced by adopting effective strategies and thus enhancing the reliability, transparency, and overall integrity of research practices.

Reporting Standards and Guides: Efforts to improve reproducibility in computational research have increasingly emphasized the importance of structured reporting, artifact documentation, and training. [43] advocated for using research Compendia to improve computational reproducibility, particularly in geography and geosciences. It calls for changes in scholarly communication and the adoption of open-source tools. [23] focused on practical skills for researchers, emphasizing the importance of organizing, documenting, and sharing code and data to enhance reproducibility in computational analysis. Together, these studies propose actionable strategies for achieving reproducibility across different scientific disciplines.

3.3 Post-Intervention Review

Some studies collectively assessed the challenges and strategies surrounding computational reproducibility in science. [12] evaluated the effectiveness of the Open Data badge in psychology, finding limited adherence to reproducibility standards, where only one article was deemed exactly reproducible. The articles published in PLOS ONE are found to have reproducibility difficulties, revealing the lack of available source code and adherence to open policies [54]. [13] examined the impact of programming workshops on biomedical researchers, noting that while participants did not show statistically significant improvements in reproducible behaviors, many adopted new practices that enhanced transparency in their workflows. [37] highlighted the importance of various research artifacts in reproducibility, advocating for comprehensive evaluations of these components.

3.4 Comparative Analysis

These studies highlight efforts to compare different articles, tools, software, and approaches in terms of reproducibility. [32] compared the reproducibility of molecular phylogenetic models, emphasizing the role of heuristic algorithms, while [28] focused on the software environment and versioning as critical factors in reproducing research outcomes. [18] examined tools for biomedical named entity recognition, showcasing how standardizing input and output formats facilitates reproducibility across different datasets. [26] compared workflows in geographic analyses, identifying both computational and conceptual barriers to reproducibility. [1] compared machine learning models and stressed the importance of comprehensive documentation and methodological transparency to achieve reproducible results. The issues in computational reproducibility are argued to stem from unrealistic expectations in computational chemistry, advocating for the need for characteristics-transparency, consistency, sustainability, and inclusivity [22].

3.5 Identifying Checklist Items from Literature Review

The existing literature on computational reproducibility in social science has two types of challenges. First, the general barriers that are common across multiple disciplines at the concept level [14, 31, 33, 43] i.e., following open science principles, licensing etc., and at implementation level [17, 19, 23, 41] i.e., using version control, containerization tools for capturing dependencies etc. Many discipline-agnostic checklist items can be easily transferred across disciplines, such as reporting the "How to Use" for using the code with original settings, freezing dependencies of the working virtual environment, and using version control tools. A specific example in this regard is the use of citation file format (CFF), which is frequently used in engineering and computational sciences as a standard to make citation metadata accessible and machine-readable while ensuring proper attribution and credit to the creators. It has been adopted in the proposed checklist, mentioned as the last item in Checklist 3 under dissemination in Figure 7. There is also a need for documentation and guides that explain the purpose and use of such cross-disciplinary checklist items to their respective audience using their terminology. The cross-disciplinary computational reproducibility checklist items are adopted from the related disciplines, while the social science-specific unique challenges are also framed as checklist items. They are all compiled into a checklist for the surveys to collect community feedback.

Computational reproducibility is gaining momentum in social sciences with recent contributions addressing concept-level barriers, i.e., defining reproducibility and related concepts [20, 53] and addressing discipline-agnostic challenges, i.e., open access and software versioning [11, 51]. At the implementation level, specialized reproducibility tools are also developed for improving the reproducibility of computational methods, e.g., Rtoot for the R language [9, 52]. Recent studies are more inclined toward social media data for analyzing societal behavioral patterns. It involves dealing with dynamically evolving data having ethical considerations and access restrictions [6]. Some of the frequently used tools in social science lack backward or cross-version compatibility. Qualitative social scientists may have less exposure to certain computational tools. Thus, despite high demand, large language models are underused in social science use cases due to inadequate documentation. Understanding the gravity of the challenges for computational reproducibility in social science, more effort focusing on formalizing data access requirements, reporting experimental details, and using open-source tools is required.

4 Survey on Reproducibility Practices and Challenges

To complement the insights from the literature, we conducted an online survey to understand how computational reproducibility is practiced and perceived within the social science research community. The survey focused on tools and practices used and the challenges faced. It received 180 responses. The majority of the participants are at senior academic positions, i.e., full professors (36%) and assistant professors (22%), with over a decade of research experience (77%), as shown in Figure 2. Although underrepresented, students, industry professionals, and other researchers also participated in the survey.

Figure 3 presents the percentage of responses on reproducibility practices and perspectives, categorized by years of experience (1–5 years, 6–10 years, and more than 10 years). It indicates that researchers with 6–10 years of experience are the most engaged with reproducibility practices, surpassing both early-career (1–5 years) and senior researchers (more than 10 years). In contrast, early-career researchers exhibit the lowest levels of engagement, suggesting potential gaps in training or awareness. These findings highlight the need for targeted initiatives to support early-career researchers adopting reproducibility practices.

Figure 4 illustrates the strategies researchers use to ensure the reproducibility of their data processing and analysis steps. The most widely adopted approach is documenting experiment steps (81%), followed by open access to code and resources (71%). Researchers particularly emphasize the importance of properly documenting data cleaning and preparation steps, as inconsistencies in these areas can lead to variations in reproducibility. However, only 29% of participants use standardized checklists and scripts for data analysis, and the adoption of specialized checklists and guides is even lower at 16%, making it the least common approach. However, this can be due to the highly experienced dominant group among the survey participants.

The barriers that hinder computational reproducibility are a lack of time, inadequate resources like checklists and guides, and a lack of awareness of the available resources. Additionally, some issues not strictly tied to computational reproducibility were also highlighted, in Figure 5. The issues identified can be summarized as follows; Computational reproducibility needs to be more obvious and its resources widely known. It must be integrated into the main research workflow in a time-efficient manner. This also involves simplifying the tools and procedures for higher adoption. There is also a need for deliberate efforts, such as workshops and training sessions, to raise awareness among early career researchers about computational reproducibility and the available resources.

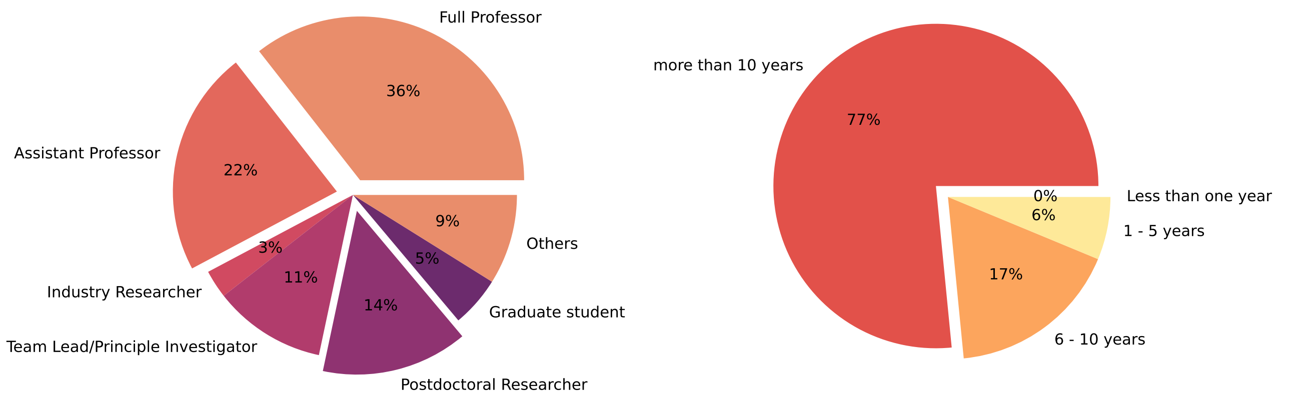
Additionally, the participants also supported three main features in the computational reproducibility efforts. First, examples of best practices should be provided to guide researchers toward the target research deliverables, as suggested by 58% of participants. Second, clear definitions and standards are supported by 56%, believing that it would significantly help in improving reproducibility. Third, to have a step-by-step guide suggested by 51%, that will make the research process transparent, in Figure 6. They preferred preserving metadata about the research covering details about the nature of the model and information about the contact person and funding body, etc. More statistics were considered worth reporting in the documentation, including assumptions, commands, and decisions. The participants also asked for more clarity on compliance with the checklists.

Most of the survey participants, i.e., 76%, affirmed the need for computational reproducibility checklists, while 91% felt the need for a portal that enforces them, Figure 3. They also suggested that in-demand computational models should be covered as guided examples on the checklist-assisted platform, such as large language models for specialized use cases, machine learning models in general, and agent-based simulations. Data access or data gathering models that acquire data from digital platforms are also in demand. They favored the interactive model documentation, e.g., Jupyter notebooks that provide explanations of settings and decisions alongside the code. There is also an urge to address the legal and ethical challenges, along with addressing negative results in research to improve the effectiveness of the checklists.

5 Survey on Checklist Items Evaluation

To evaluate the perceived usefulness of the identified checklist items, the second survey received 64 responses to 59 checklist items. They are rated on a 9-point Likert scale from 1 as strongly disagree to 9 as strongly agree. Percentages of responses to the checklist items are presented in Table 2 in the Appendix. For practical purposes, we group the responses as **7 to 9**, indicating highly demanded, i.e., to be a mandatory checklist item, **4 to 6** as neutral, i.e., optional checklist item, and **1 to 3** as exclusion, i.e., to be excluded from the final checklist. Some of the checklist items were skipped as well by the participants. The values represent percentages for a specific checklist item, e.g., "Table2-Q1# Hypothesis: Clearly state the hypotheses" is decided inclusion by 82.81% participants, exclusion by 7.81% participants, neutral by 7.81%, and is skipped by 1.56% participants.

Key Trends from the Survey: The survey results reveal a strong tendency towards supporting the inclusion of many methodological aspects in research documentation. Several questions received exceptionally high agreement as inclusion, indicating a broad consensus on their importance. Notably, "Table2-Q26# Statistical and Computational Methods: Describe all analytical methods" has the highest inclusion of 96.88%, while "Table2-Q4# Objective: Describe the study objectives", "Table2-Q15# Data Sources", and "Table2-Q16# Data Description" received 95.31% inclusion. These



(a) Percentage of survey participants with certain positions **(b) Percentage of survey participants with certain years of experience**

Figure 2: Demographics of the survey participants. The majority of the participants are in top positions (full and assistant professor; on the left) and have more than 10 years of experience (on the right).

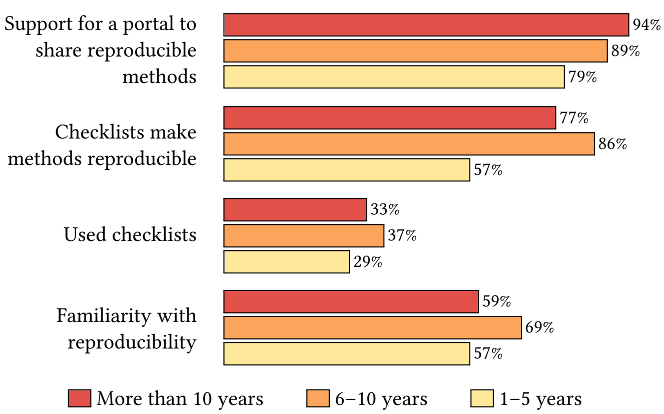


Figure 3: Percentage of participants selecting specific responses on reproducibility practices and perspectives, for each experience group (1–5 years, 6–10 years, and more than 10 years of experience).

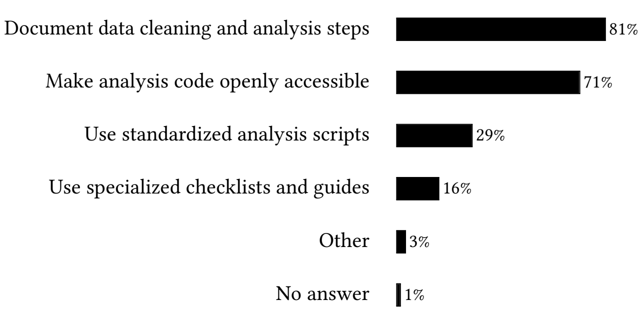


Figure 4: Percentage of participants using different approaches to enhance computational reproducibility.

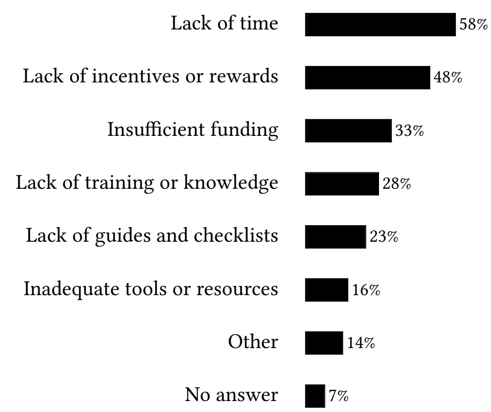


Figure 5: Challenges hindering computational reproducibility in social science.

findings suggest that participants strongly value comprehensive methodological transparency, particularly regarding statistical techniques, data characterization, and study objectives. This trend likely stems from the increasing emphasis on reproducibility, accountability, and structured methodological approaches in contemporary research.

Conversely, certain methodological elements saw notable agreement for exclusion. The most frequently excluded aspects included "Table2-Q31# Training and Course Availability" with 28.13% exclusion, while "Table2-Q45# Showing Sample Input and Output Data" and "Table2-Q46# Providing Step-by-Step How-Tos" have 21.88% exclusion, while "Table2-Q39# Including social science Use Cases " has 18.75% exclusion. The higher exclusion rates in these categories suggest that respondents may perceive these elements as supplementary rather than essential components of methodological documentation. Potential reasons for these responses could be practical limitations in documenting every instructional aspect, as

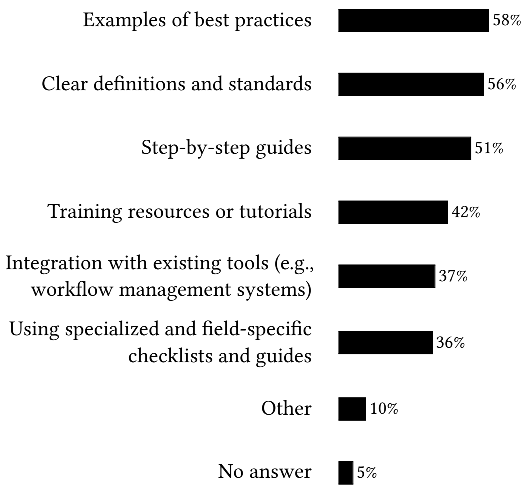


Figure 6: Key features to overcome the computational reproducibility challenges.

well as a distinction between methodological reporting and educational resources. Additionally, use-case documentation may be considered context-dependent, with researchers possibly preferring case studies to be presented separately from core methodological descriptions. The checklist items with lower demand address the needs of early career researchers who are underrepresented in the survey. They were also more unaware and lesser keen to make their work reproducible due to a lack of resources and time. These findings indicate prioritization of core methodological clarity over supplementary instructional materials, reflecting a focus on essential research documentation.

6 Proposed Computational Reproducibility Checklists for Social Sciences

In this section, we propose checklists for computational reproducibility in social science as an extension of the TIER2 Pilot 4 Reproducibility checklist[4]. They are designed to improve the longevity and usability of computational social science research artifacts consisting of items with highest inclusion percentages from the checklist items evaluation Survey or based on the conclusions drawn from the practices and open challenges Survey. The checklists capture the specific needs of computational reproducibility in social science research, providing practical and understandable directives. It will be preserved on a long-term archival platform and improved in subsequent versions.

Figure 7 presents the proposed checklists inspired by four of the 5 pillars of reproducibility for bioinformatics, focusing on code versioning, persistent sharing, virtual environment preservation, and documentation [63]. The code versioning and environment preservation are supplemented with coding conventions for readability, understandability, and adaptation. Data access and preservation are addressed in Checklist 1, while sharing and archiving are addressed through Checklist 3. The checklist items are organized into three checklists based on the research stages for easy adoption alongside the research process. Although it specifically focuses on social science, other disciplines may benefit from it in different aspects of computational reproducibility.

7 Discussion and Future Work

Our proposed checklists provide a structured yet flexible solution to help operationalize abstract principles of reproducibility in a practical, actionable, and discipline-specific manner. Based on community feedback that we collected, the checklists address the need for computational reproducibility in social science. Our first survey investigated reproducibility practices and challenges, revealing considerably lower adoption of specialized checklists and guides for computational reproducibility in social sciences compared to other areas of concern. In related disciplines, like psychology and political science, a higher adoption of such resources resulted in improved computational reproducibility. Therefore, to increase adoption in social science, we aim for checklists that are more practical, actionable, clear, and less labor-intensive. As part of this endeavor, we also provide examples that adhere to the checklists. Moreover, to meet the needs of early-career researchers, additional support and details are not included in the checklists but rather references to it for further support. Moreover, to lower the typical effort necessary for compliance, we aligned the checklists with preferred tools in the community, e.g., using Jupyter Notebooks, GitHub repositories, and containerized environments. The second survey provides feedback on specific checklist items to consider, including and refining them further for clarity and purpose. However, we acknowledge that our sample is not representative of the broader social sciences research community. As such, the findings should be interpreted as exploratory rather than conclusive. Nonetheless, the results offer meaningful insights into current practices and gaps, which informed the design and prioritization of the checklist items.

The proposed checklists for computational reproducibility of social science models share similarities with existing checklists. For example, the FAIRER aware reproducibility checklist[5] reflects similar technical considerations with our proposed checklists. However, the two checklists differ in two important ways. First, the proposed checklists are solely focused on computational reproducibility and therefore exclude items related to methodology, results, or interpretative claims. Second, this narrow focus allows for a more detailed treatment of the technical and reporting dimensions of computational reproducibility, making the checklists more actionable and aligned with practical adoption.

To meet the expectations of early career researchers without overburdening experienced researchers with excessive details, the checklists are complemented with detailed computational reproducibility guides. The checklist items directly link to the specific sections on the guides that explain them and provide actionable instructions to meet the requirements. The guides assist with further exploration, explanation, technical assistance, and possible alternatives, while also providing background for making certain decisions. Thus, the guides provide further help in complying with the checklist. The guides state the scope of the checklists up front,

[4]https://tier2-project.eu/news/tier2-pilot-4-enhance-reproducibility-code-and-data-computational-social-science

[5]https://fairerdata.github.io/FAIRER-Aware-REPRODUCIBILITY-Assessment/

Checklist 1: Data access methods (planning and data collection)

- **Data access**
  - ☐ *Ethical restrictions:* The data collection tools and methods have direct user consent and do not store personal information
  - ☐ *Documentation:* Documenting the data collection process, i.e., API used, date, and configuration settings, while complying with the terms of service
  - ☐ *Security:* Using data encryption methods to ensure the data is stored securely for the time of experimentation and properly disposed of.
  - ☐ *Validation:* Validation methods that check for correctness and completeness of data
- **Planning**
  - ☐ *Planning sources:* Planning for methods to correct data for relevant sources through APIs or scraping
  - ☐ *Deciding analysis model:* Choosing the analysis model from the openly available models, relevant to the study and its data
  - ☐ *Sampling Strategy:* Defining methods that evaluate the usefulness of data for the study as selection criteria

Checklist 2: Analysis methods

- **Readability and understandability**
  - ☐ *Code:* Follow basic coding conventions, while making good use of comments and white spaces[a]
  - ☐ *Documentation:* Documenting the research setup and model configurations[b]
  - ☐ *Version:* Using version control tool e.g., Git.
  - ☐ *Using reproducibility tools:* Deploy software setup to isolate and preserve the research environment, e.g., dockerizing it
- **Ease of reuse in code execution**
  - ☐ *Commands:* Maintaining commands log to recreate the setup
  - ☐ *Code execution:* An-easy-to-follow *"How to Use"* that reproduce results on sample data even for non-technical users
  - ☐ Providing sample input and output data to replicate for proof of concept

Checklist 3: Sharing and archiving procedures

- **Making all resources available**
  - ☐ *Code:* Sharing the code as a public repository, e.g., on GitHub. Ideally, a Digital Object Identifier (DOI) is assigned to the working version of the code for persistent sharing
  - ☐ *Documentation:* The documentation, e.g., a well-written README, should be made part of the repository. The README should provide all necessary details to recreate the environment and reproduce results from the experiment[c]
  - ☐ *Preserving the working environment:* Preserving the working environment of the method, i.e., required libraries, packages, and their version, e.g., by generating *requirements.txt*
  - ☐ *Data:* Making the research data collection process and the data handling public while staying within ethical boundaries for sensitive information
- **Accessibility**
  - ☐ *Public availability:* All resources used in the experiment are open-source and publicly available
  - ☐ *Provided on request:* Sensitive information needed to recreate the study is provided on request as an explicit message. In case of sharing from personal/organizational pages, ensure the link is active and accessible
  - ☐ Integrating the research resources with execution environment or ensuring their access through public development environments, e.g., MyBinder[d] for easy and quick proof of concept
- **Licensing**
  - ☐ *Types of licenses:* A license must be added to the repository. The commonly used open licenses on GitHub are MIT, Apache 2.0, and CC By 4.0
  - ☐ *Openness of licenses:* The licenses allow different levels of reuse of the existing work. However, ideally, everything should be free for any kind of reuse
- **Dissemination**
  - ☐ Demonstrating the use of the method through a step-by-step guide as a tutorial[e]
  - ☐ Having a citation file to help in citing the method[f]

[a]https://github.com/GESIS-Methods-Hub/guidelines/tree/v0?tab=readme-ov-file#quality-criteria
[b]https://github.com/GESIS-Methods-Hub/guidelines/blob/v0/method/template.md
[c]https://github.com/GESIS-Methods-Hub/guidelines
[d]https://mybinder.org/
[e]https://github.com/GESIS-Methods-Hub/guidelines/tree/v0/tutorial
[f]https://citation-file-format.github.io/

Figure 7: Checklists for computational reproducibility in social sciences, updated and refined according to survey feedback.

mentioning their application for computational research in social science only. However, in parts, the checklists may also benefit computational reproducibility in other disciplines. We also complemented the checklists with reporting templates to assist with structured, standardized, and consistent reporting across computational methods. The templates contribute to uniformity in reporting while minimizing human effort and chances of making errors. For example, the checklist requires a "How to Use" section in the model report. The guides explain it and provide information to prepare it. The template specifications instruct its reporting format, While the checklist-compliant samples practically demonstrate it. With this high level of clarity and support, we want to encourage more researchers to comply with the reproducibility standards, improving overall research credibility [21, 42]. The guides, templates, and demonstrative examples were not provided along with the checklist items in the surveys, but will be part of future studies.

Providing the checklists in itself does not guarantee adherence, but rather facilitates it. Through explicit expectations, the checklists serve both as a tool for self-assessment during the research process and as a reference framework for external reviewers. Our next goal is to integrate the complementary material of the checklist, i.e., guides, templates, and demonstrative examples, into everyday research workflows. We realize this through the *Methods Hub*[6] portal, an infrastructure for finding, accessing, understanding, sharing, and executing reproducible computational models for social sciences. The portal uses the proposed checklists as inclusion criteria, supplemented through guides and templates for further assistance. It is aligned with the FAIR principles and reproducibility practices. It encourages the computing and computational social science community to develop models complying with the proposed checklists and share them through the portal, offering multiple incentives. the portal supports a journal-like review mechanism where the submitted model is kept in moderation until the deviations from the checklists are addressed. Only then is the model published on the portal. Future studies on the portal will evaluate improvement in computational reproducibility through the adoption of the checklists. Comparative replication studies using models complying with the checklists and published on the portal with their alternative implementations are also planned in the future.

Interestingly, while respondents supported structured practices in principle, they perceived items such as "instructional", i.e., tutorials or step-by-step examples, to be excessive for the checklists. However, the early career researchers, underrepresented in the survey, find them helpful. We thus also plan to develop modular training materials to support adoption among early-career researchers who may lack formal exposure to reproducibility tools. This indicates a distinction between what researchers view as essential to reproduce a study and what they consider supportive but optional. It may also reflect a tradeoff between methodological rigor and perceived constraints on creativity or time. Our results also suggest a gap in engagement with reproducibility practices among early-career researchers, who may lack exposure to reproducibility tools and training opportunities. The survey data suggests that targeted interventions—such as mentorship programs, reproducibility-focused workshops, and the integration of structured checklists into graduate curricula—could help bridge this gap. Empowering emerging researchers with the knowledge and resources to embed reproducibility into their workflow from the outset could lead to a cultural shift toward more transparent and robust research practices.

[6]https://methodshub.gesis.org/

8 Conclusion

In this research, we discussed existing literature on computational reproducibility in social science and related disciplines. The systematic literature review presents the current state of computational reproducibility in social science and related disciplines. It also highlighted the specialized needs of computational reproducibility, such as dealing with data access requirements. We identify 59 checklist items from the resources available in the literature. Two surveys are conducted to gather feedback on the practices followed and challenges faced by the community, while also evaluating the necessity of the identified checklist items. Observing inclusion percentages on the checklist items, only mandatory checklist items are retained in the final checklists. The checklist items are organized into three checklists aligned with the research process. The survey on the proposed checklists has received wider acceptance by 98.43% of users, considering 76.35% of items crucial while others useful. The survey participants also strongly agreed (above 95%) to have data description and source documented in the experiment. The checklists reflect the computational reproducibility needs of the social science models; however, the participants in surveys may not be representative of the social science community in general. The extension of this work includes computational reproducibility guides, reporting templates, a portal to publish models complying with the checklists, and a journal-like review mechanism to not only guide but ensure compliance with the checklist.

Data and Code Availability

All materials, including the survey instruments, anonymized response data, and checklist resources, are published in the associated public repository[7].

Acknowledgments

This research was supported by TIER2 (Grant No. 101094817), with additional support from GESIS – Leibniz Institute for the Social Sciences through the Digital Behavioral Data project. We also acknowledge the German Competence Centre for Bibliometrics for providing access to the in-house Scopus database.

References

[1] Akhil Pandey Akella, David Koop, and Hamed Alhoori. 2023. Laying Foundations to Quantify the 'Effort of Reproducibility'. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. IEEE Computer Society, Los Alamitos, CA, USA, 56–60. doi:10.1109/jcdl57899.2023.00018

[2] Adam Altmejd, Anna Dreber, Eskil Forsell, Juergen Huber, Taisuke Imai, Magnus Johannesson, Michael Kirchler, Gideon Nave, and Colin Camerer. 2019. Predicting the replicability of social science lab experiments. *PloS one* 14, 12 (2019), e0225826.

[3] Althea A. Archmiller, Andrew D. Johnson, Jane Nolan, Margaret Edwards, Lisa H. Elliott, Jake M. Ferguson, Fabiola Iannarilli, Juliana Vélez, Kelsey Vitense, Douglas H. Johnson, and John Fieberg. 2020. Computational Reproducibility in The Wildlife Society's Flagship Journals. *Journal of Wildlife Management* 84 (2020), 1012–1017. doi:10.1002/jwmg.21855

[4] Richard Ball, Norm Medeiros, Nicholas W. Bussberg, and Aneta Piekut. 2022. An Invitation to Teaching Reproducible Research: Lessons from a Symposium. *Journal of Statistics and Data Science Education* 30 (2022), 209–218. doi:10.1080/26939169.2022.2099489

[5] Soudabeh Barghi, Lalet Scaria, Ali Salari, and Tristan Glatard. 2018. Predicting computational reproducibility of data analysis pipelines in large population studies using collaborative filtering. In *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*. IEEE, Seattle, WA, USA, 8622095–8622095. doi:10.1109/bigdata.2018.8622095

[7]https://github.com/momenifi/reproducibility_checklist_CSS

[6] Marco Bastos. 2021. This account doesn't exist: Tweet decay and the politics of deletion in the Brexit debate. *American Behavioral Scientist* 65, 5 (2021), 757–773.

[7] Asti Bhatt, Todd Valentic, Ashton Reimer, Leslie Lamarche, Pablo Reyes, and Russell Cosgrove. 2020. Reproducible Software Environment: A tool enabling computational reproducibility in geospace sciences and facilitating collaboration. *Journal of Space Weather and Space Climate* 10 (2020), swsc190063. doi:10.1051/swsc/2020011

[8] Tristan Campbell, Kingsley W. Dixon, and Rebecca N. Handcock. 2023. Restoration and replication: a case study on the value of computational reproducibility assessment. *Restoration Ecology* 31 (2023), 1–9. Issue 8. doi:10.1111/rec.13968

[9] Chung-hong Chan and David Schoch. 2023. rang: Reconstructing reproducible R computational environments. *Plos one* 18, 6 (2023), e0286761.

[10] Kyle Chard, Niall Gaffney, Matthew B. Jones, Kacper Kowalik, Bertram Ludäscher, Jarek Nabrzyski, Victoria Stodden, Ian Taylor, Matthew J. Turk, and Craig Willis. 2019. Implementing computational reproducibility in the Whole Tale environment. In *P-RECS 2019 - Proceedings of the 2nd International Workshop on Practical Reproducible Evaluation of Computer Systems, co-located with HPDC 2019*. Association for Computing Machinery, New York, NY, USA, 17–22. doi:10.1145/3322790.3330594

[11] April Clyburne-Sherin, Xu Fei, and Seth Ariel Green. 2019. Computational reproducibility via containers in psychology. *Meta-psychology* 3 (2019), 1–9.

[12] Sophia Crüwell, Deborah Apthorp, Bradley J. Baker, Lincoln Colling, Malte Elson, Sandra J. Geiger, Sebastian Lobentanzer, Jean Monéger, Alex Patterson, D. Samuel Schwarzkopf, and Mirela Zaneva. 2023. What's in a Badge? A Computational Reproducibility Investigation of the Open Data Badge Policy in One Issue of Psychological Science. *Psychological Science* 34 (2023), 512–522. doi:10.1177/09567976221140828

[13] Ariel Deardorff. 2020. Assessing the impact of introductory programming workshops on the computational reproducibility of biomedical workflows. *PLoS ONE* 15, e0230697 (2020), 1–11. doi:10.1371/journal.pone.0230697

[14] Mine Dogucu and Mine Çetinkaya Rundel. 2022. Tools and Recommendations for Reproducible Teaching. *Journal of Statistics and Data Science Education* 30 (2022), 251–260. doi:10.1080/26939169.2022.2138645

[15] Bakinam T Essawy, Jonathan L Goodall, Wesley Zell, Daniel Voce, Mohamed M Morsy, Jeffrey Sadler, Zhihao Yuan, and Tanu Malik. 2018. Integrating scientific cyberinfrastructures to improve reproducibility in computational hydrology: Example for HydroShare and GeoTrust. *Environmental Modelling and Software* 105 (2018), 217–229.

[16] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. 2019. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM conference on recommender systems*. Association for Computing Machinery, New York, NY, USA, 101–109.

[17] Luiz Gadelha, Martin Hohmuth, Mahnoor Zulfiqar, David Schone, Sheeba Samuel, Maria Sorokina, Christoph Steinbeck, and Birgitta Konig-Ries. 2022. Toward a Framework for Integrative, FAIR, and Reproducible Management of Data on the Dynamic Balance of Microbial Communities. In *Proceedings - 2022 IEEE 18th International Conference on e-Science, eScience 2022*. IEEE, Salt Lake City, UT, USA, 443–449. doi:10.1109/escience55777.2022.00080

[18] Ana Garcia-Serrano, Sebastian Hennig, and Andreas Nürnberger. 2021. Computational reproducibility of named entity recognition methods in the biomedical domain. *Procesamiento del Lenguaje Natural* 66 (2021), 141–152. doi:10.26342/2021-66-12

[19] Michael R Gryk and Bertram Ludäscher. 2018. Semantic mediation to improve reproducibility for biomolecular NMR analysis. In *Lecture Notes in Computer Science*, Vol. 10766. Springer, Springer, Cham, Cham, 620–625.

[20] Odd Erik Gundersen. 2021. The fundamental principles of reproducibility. *Philosophical Transactions of the Royal Society A* 379, 2197 (2021), 20200210.

[21] Tom E Hardwicke, Joshua D Wallach, Mallory C Kidwell, Theiss Bendixen, Sophia Crüwell, and John PA Ioannidis. 2020. An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *Royal Society Open Science* 7, 2 (2020), 190806.

[22] Alexandre Hocquet and Frédéric Wieber. 2021. Epistemic issues in computational reproducibility: software as the elephant in the room. *European Journal for Philosophy of Science* 11 (2021), 38. doi:10.1007/s13194-021-00362-9

[23] William A. Ingram and Edward A. Fox. 2020. Preparing code and data for computational reproducibility. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. ACM, Association for Computing Machinery, New York, NY, USA, 565–566. doi:10.1145/3383583.3398714

[24] Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T Evelo, et al. 2020. FAIR principles: interpretations and implementation considerations. 10–29 pages.

[25] Kwangwoog Jung, Yang-Ki Cho, and Yong-Jin Tak. 2021. Containers and orchestration of numerical ocean model for computational reproducibility and portability in public and private clouds: Application of ROMS 3.6. *Simulation Modelling Practice and Theory* 109 (2021), 102305. doi:10.1016/j.simpat.2021.102305

[26] Peter Kedron, Sarah Bardin, Joseph Holler, Joshua Gilman, Bryant Grady, Megan Seeley, Xin Wang, and Wenxin Yang. 2023. A Framework for Moving Beyond Computational Reproducibility: Lessons from Three Reproductions of Geographical Analyses of COVID-19. *Geographical Analysis* 56 (2023), 163–184. Issue 1. doi:10.1111/gean.12370

[27] Peter Kedron, Wenwen Li, Stewart Fotheringham, and Michael Goodchild. 2021. Reproducibility and replicability: Opportunities and challenges for geospatial research. *International Journal of Geographical Information Science* 35, 3 (2021), 427–445.

[28] Jenna Kim, Catherine Blake, and Peter T. Darch. 2019. Toward computational reproducibility: A doctoral student's story of passing the baton. *Proceedings of the Association for Information Science and Technology* 56 (2019), 688–690. doi:10.1002/pra2.134

[29] Philipp Knöpfle, Mario Haim, and Johannes Breuer. 2024. Key topic or bare necessity? How Research Ethics are Addressed and Discussed in Computational Communication Science. *Publizistik* 69, 3 (2024), 333–356.

[30] Florian Koenigstorfer, Armin Haberl, Dominik Kowald, Tony Ross-Hellauer, and Stefan Thalmann. 2024. Black Box or Open Science? Assessing Reproducibility-Related Documentation in AI Research. In *Hawaii International Conference on System Sciences 2024 (HICSS-57)*. University of Hawaii at Manoa, Hilton Hawaiian Village, Honolulu, Hawaii, 682–691.

[31] Markus Konkol, Christian Kray, and Max Pfeiffer. 2019. Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. *International Journal of Geographical Information Science* 33 (2019), 408–429.

[32] Sudhir Kumar, Qiqing Tao, Alessandra P. Lamarca, and Koichiro Tamura. 2023. Computational Reproducibility of Molecular Phylogenies. *Molecular Biology and Evolution* 40 (2023), msad165. doi:10.1093/molbev/msad165

[33] Nicole A. Lazar. 2023. Functional neuroimaging in the era of Big Data and Open Science: A modern overview. *Wiley Interdisciplinary Reviews: Computational Statistics* 15 (2023), e1609. doi:10.1002/wics.1609

[34] David MJ Lazer, Alex Pentland, Duncan J Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts, et al. 2020. Computational social science: Obstacles and opportunities. *Science* 369, 6507 (2020), 1060–1062.

[35] Sabina Leonelli. 2018. Rethinking reproducibility as a criterion for research quality. In *Including a symposium on Mary Morgan: curiosity, imagination, and surprise*, Vol. 36. Emerald Publishing Limited, Leeds, UK, 129–146.

[36] David M. Liu and Matthew J. Salganik. 2019. Successes and Struggles with Computational Reproducibility: Lessons from the Fragile Families Challenge. *Socius* 5 (2019), 2378023119849803. doi:10.1177/2378023119849803

[37] Tanu Malik, Anjo Vahldiek-Oberwagner, Ivo Jimenez, and Carlos Maltzahn. 2022. Expanding the Scope of Artifact Evaluation at HPC Conferences: Experience of SC21. In *P-RECS 2022 - Proceedings of the 5th International Workshop on Practical Reproducible Evaluation of Computer Systems, co-located with HPDC 2022*. Association for Computing Machinery, New York, NY, USA, 3–9. doi:10.1145/3526062.3536354

[38] Alexander Michels, Anand Padmanabhan, Zhiyu Li, and Shaowen Wang. 2023. EasyScienceGateway: A new framework for providing reproducible user environments on science gateways. *Concurrency and Computation: Practice and Experience* 36 (2023), 1–16. Issue 4. doi:10.1002/cpe.7929

[39] K. Jarrod Millman, Satra Ghosh, and Kamil S. Jaron. 2018. Teaching Computational Reproducibility for Neuroimaging. *Frontiers in Neuroinformatics* 12 (2018), 72. doi:10.3389/fninf.2018.00072

[40] James W Moody, Lisa A Keister, and Maria C Ramos. 2022. Reproducibility in the social sciences. *Annual review of sociology* 48, 1 (2022), 65–85.

[41] David Moreau, Kristina Wiebels, and Carl Boettiger. 2023. Containers for computational reproducibility. *Nature Reviews Methods Primers* 3 (2023), 50. doi:10.1038/s43586-023-00236-9

[42] Brian A Nosek, George Alter, George C Banks, Denny Borsboom, Sara D Bowman, Steven J Breckler, Stuart Buck, Christopher D Chambers, Gilbert Chin, Garret Christensen, et al. 2015. Promoting an open research culture. *Science* 348, 6242 (2015), 1422–1425.

[43] Daniel Nüst and Edzer Pebesma. 2021. Practical Reproducibility in Geography and Geosciences. *Annals of the American Association of Geographers* 111 (2021), 1300–1310. doi:10.1080/24694452.2020.1806028

[44] Pepijn Obels, Daniël Lakens, Nicholas A. Coles, Jaroslav Gottfried, and Seth A. Green. 2020. Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology. *Advances in Methods and Practices in Psychological Science* 3 (2020), 229–237. doi:10.1177/2515245920918872

[45] Thomas Pasquier, Matthew K Lau, Xueyuan Han, Elizabeth Fong, Barbara S Lerner, Emery R Boose, Merce Crosas, Aaron M Ellison, and Margo Seltzer. 2018. Sharing and preserving computational analyses for posterity with encapsulator. *Computing in Science and Engineering* 20 (2018), 111–124.

[46] Limor Peer, Lilla V. Orr, and Alexander Coppock. 2021. Active Maintenance: A Proposal for the Long-Term Computational Reproducibility of Scientific Results. *Frontiers in Research Metrics and Analytics* 6 (2021), 639178. doi:10.3389/frma.2021.639178

[47] Michel Tuan Pham and Travis Tae Oh. 2021. Preregistration is neither sufficient nor necessary for good science. *Journal of Consumer Psychology* 31, 1 (2021),

163–176.

[48] Line Pouchard, Sterling Baldwin, Todd Elsethagen, Shantenu Jha, Bibi Raju, Eric Stephan, Li Tang, and Kerstin Kleese Van Dam. 2019. Computational reproducibility of scientific workflows at extreme scales. *International Journal of High Performance Computing Applications* 33 (2019), 763–776. doi:10.1177/1094342019839124

[49] Ali Salari, Yohan Chatelain, Gregory Kiar, and Tristan Glatard. 2021. Accurate Simulation of Operating System Updates in Neuroimaging Using Monte-Carlo Arithmetic. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12959 LNCS (2021), 14–23. doi:10.1007/978-3-030-87735-4_2

[50] Lamia Salsabil, Jian Wu, Muntabir Hasan Choudhury, William A. Ingram, Edward A. Fox, Sarah M. Rajtmajer, and C. Lee Giles. 2022. A Study of Computational Reproducibility using URLs Linking to Open Access Datasets and Software. In *WWW 2022 - Companion Proceedings of the Web Conference 2022*. Association for Computing Machinery, New York, NY, USA, 784–788. doi:10.1145/3487553.3524658

[51] Sheeba Samuel and Daniel Mietchen. 2024. Computational reproducibility of Jupyter notebooks from biomedical publications. *GigaScience* 13 (2024), giad113.

[52] David Schoch and Chung-hong Chan. 2023. Software presentation: Rtoot: Collecting and Analyzing Mastodon Data. *Mobile Media & Communication* 11 (2023), 575–578. Issue 3.

[53] David Schoch, Chung-hong Chan, Claudia Wagner, and Arnim Bleier. 2024. Computational Reproducibility in Computational Social Science. *EPJ Data Science* 13, 1 (2024), 1–11.

[54] Heidi Seibold, Severin Czerny, Siona Decke, Roman Dieterle, Thomas Eder, Steffen Fohr, Nico Hahn, Rabea Hartmann, Christoph Heindl, Philipp Kopper, Dario Lepke, Verena Loidl, Maximilian Mandl, Sarah Musiol, Jessica Peter, Alexander Piehler, Elio Rojas, Stefanie Schmid, Hannah Schmidt, Melissa Schmoll, Lennart Schneider, Xiao-Yin To, Viet Tran, Antje Völker, Moritz Wagner, Joshua Wagner, Maria Waize, Hannah Wecker, Rui Yang, Simone Zellner, and Malte Nalenz. 2021. A computational reproducibility study of PLOS ONE articles featuring longitudinal data analyses. *PLoS ONE* 16, NULL (2021), e0251194. doi:10.1371/journal.pone.0251194

[55] Heidi Seibold, Severin Czerny, Siona Decke, Roman Dieterle, Thomas Eder, Steffen Fohr, Nico Hahn, Rabea Hartmann, Christoph Heindl, Philipp Kopper, Dario Lepke, Verena Loidl, Maximilian Mandl, Sarah Musiol, Jessica Peter, Alexander Piehler, Elio Rojas, Stefanie Schmid, Hannah Schmidt, Melissa Schmoll, Lennart Schneider, Xiao-Yin To, Viet Tran, Antje Völker, Moritz Wagner, Joshua Wagner, Maria Waize, Hannah Wecker, Rui Yang, Simone Zellner, and Malte Nalenz. 2022. Erratum: A computational reproducibility study of PLOS ONE articles featuring longitudinal data analyses. *PLoS ONE* 17 (2022), 1–15. doi:10.1371/journal.pone.0269047

[56] Harald Semmelrock, Simone Kopeinik, Dieter Theiler, Tony Ross-Hellauer, and Dominik Kowald. 2023. Reproducibility in Machine Learning-Driven Research. *arXiv preprint arXiv:2307.10320* (2023), 1–15.

[57] Joseph Shenouda and Waheed U. Bajwa. 2023. A Guide to Computational Reproducibility in Signal Processing and Machine Learning [Tips & Tricks]. *IEEE Signal Processing Magazine* 40 (2023), 141–151. doi:10.1109/msp.2022.3217659

[58] Fábio Silva and Cesar Analide. 2019. Computational sustainability and the PHESS platform: Using affective computing as social indicators. *Future Generation Computer Systems* 92 (2019), 329–341.

[59] Adam H. Sparks, Emerson M. Del Ponte, Kaique S. Alves, Zachary S.L. Foster, and Niklaus J. Grünwald. 2023. Openness and Computational Reproducibility in Plant Pathology: Where We Stand and a Way Forward. *Phytopathology* 113 (2023), 1159–1170. doi:10.1094/phyto-10-21-0430-per

[60] Victoria Stodden, Jennifer Seiler, and Zhaokun Ma. 2018. An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences of the United States of America* 115 (2018), 2584–2589.

[61] Ana Trisovic, Philip Durbin, Tania Schlatter, Gustavo Durand, Sonia Barbosa, Danny Brooke, and Mercè Crosas. 2020. Advancing Computational Reproducibility in the Dataverse Data Repository Platform. In *P-RECS 2020 - Proceedings of the 3rd International Workshop on Practical Reproducible Evaluation of Computer Systems*. ACM, Association for Computing Machinery, New York, NY, USA, 15–20. doi:10.1145/3391800.3398173

[62] Dandong Yin, Yan Liu, Hao Hu, Jeff Terstriep, Xingchen Hong, Anand Padmanabhan, and Shaowen Wang. 2019. CyberGIS-Jupyter for reproducible and scalable geospatial analytics. *Concurrency and Computation: Practice and Experience* 31, 4 (2019), e5040.

[63] Mark Ziemann, Pierre Poulain, and Anusuiya Bora. 2023. The five pillars of computational reproducibility: bioinformatics and beyond. *Briefings in Bioinformatics* 24, 6 (2023), bbad375.

[64] Arkaitz Zubiaga. 2018. A longitudinal assessment of the persistence of twitter datasets. *Journal of the Association for Information Science and Technology* 69, 8 (2018), 974–984.

Appendix

Table 2: Survey Results on Reproducibility Practices and Perspectives

Question	Include (%)	Exclude (%)	Neutral (%)	Skipped (%)
Q1: Hypotheses: Clearly state the hypotheses	82.81	7.81	7.81	1.56
Q2: Rationale and Prior Evidence: Summarize the study rationale and prior evidence	82.81	7.81	9.38	0.00
Q3: Study Questions: Formulate the study questions	87.50	4.69	6.25	1.56
Q4: Objectives: Describe the study objectives	95.31	3.13	1.56	0.00
Q5: Pre-registration: Register the study, including a detailed experiment design	42.19	17.19	34.38	6.25
Q6: Data Management Plan: Document the plan prior to study initiation	56.25	17.19	23.44	3.13
Q7: Statistical Analysis Plan: Outline the planned statistical procedures	82.81	6.25	9.38	1.56
Q8: Reporting Intent: Specify if methods are intended to be reported in a study	78.13	4.69	7.81	9.38
Q9: Registered Report: Indicate if the study will be submitted as a registered report	65.63	6.25	14.06	14.06
Q10: Define demographic characteristics, geographic location, and contextual factors	89.06	4.69	3.13	3.13
Q11: Sampling Strategy: Explain sampling strategy and justify the approach	89.06	1.56	6.25	3.13
Q12: Data Sources and Inclusion/Exclusion Criteria	93.75	1.56	1.56	3.13
Q13: Ethical Considerations: Address ethical considerations regarding privacy and consent	67.19	12.50	12.50	7.81
Q14: Relevance to Research Questions: Discuss relevance and comparative analysis	67.19	10.94	14.06	7.81
Q15: Data Sources: Specify digital data sources and provide DOI if available	95.31	0.00	3.13	1.56
Q16: Data Description: Briefly describe data and metadata	95.31	1.56	3.13	0.00
Q17: Accessibility: Where is the data planned to be stored?	73.44	7.81	15.63	3.13
Q18: Data Collection Procedures: Detail procedures (e.g., APIs, web scraping)	89.06	0.00	4.69	6.25
Q19: Data Pre-processing: Describe cleaning and transformation methods	92.19	0.00	3.13	4.69
Q20: Data Handling Procedures: Address handling of missing data and outliers	90.63	1.56	3.13	4.69
Q21: Computational Environment: Specify tools, software, and languages	79.69	6.25	9.38	4.69
Q22: Experimental Steps: Outline the procedure followed	84.38	1.56	6.25	7.81
Q23: Describe experimental variables	84.38	3.13	6.25	6.25
Q24: Data Collection Bias: Approaches to mitigate bias	82.81	3.13	6.25	7.81
Q25: Metadata: Provide accessible metadata	60.94	12.50	14.06	12.50
Q26: Statistical and Computational Methods: Describe all analytical methods	96.88	0.00	3.13	0.00
Q27: Algorithm and Model Description	85.94	3.13	10.94	0.00
Q28: Model Validation and Evaluation	82.81	7.81	9.38	0.00
Q29: Computational Tools and Libraries	78.13	9.38	12.50	0.00
Q30: Parameter Tuning and Optimization	84.38	4.69	7.81	3.13
Q31: Training and Course Availability	43.75	28.13	25.00	3.13
Q32: Configuration and Reproducibility	81.25	9.38	7.81	1.56
Q33: Data Handling Procedures	90.63	1.56	7.81	0.00
Q34: Interpretation of Results	87.50	7.81	4.69	0.00
Q35: Results Presentation	82.81	12.50	3.13	1.56
Q36: Code Availability	90.63	7.81	1.56	0.00
Q37: Documentation, literate programming principles	70.31	7.81	9.38	12.50
Q38: Follow existing documentation templates	62.50	10.94	14.06	12.50
Q39: Include social science use cases	43.75	18.75	28.13	9.38
Q40: Cite supporting social science literature	53.13	17.19	20.31	9.38
Q41: Explain method structure	73.44	4.69	7.81	14.06
Q42: Environment setup instructions	70.31	7.81	9.38	12.50
Q43: List required packages and versions	76.56	7.81	6.25	9.38
Q44: Specify hardware requirements	60.94	10.94	17.19	10.94
Q45: Show sample input data and corresponding output	60.94	21.88	12.50	4.69
Q46: Provide step-by-step How Tos	57.81	21.88	12.50	7.81
Q47: Include installation/configuration instructions	53.13	20.31	20.31	6.25
Q48: Deviation tracking	78.13	10.94	7.81	3.13
Q49: Report negative results	78.13	15.63	4.69	1.56
Q50: Transparent reporting of deviations	76.56	12.50	10.94	0.00
Q51: Results reporting	92.19	1.56	3.13	3.13
Q52: Interpret results w.r.t. objectives	87.50	6.25	3.13	3.13
Q53: Provide data/result visualizations	73.44	10.94	12.50	3.13
Q54: Discuss strengths and limitations	82.81	7.81	6.25	3.13
Q55: Author/contributor roles	71.88	14.06	10.94	3.13
Q56: Processed Data: Make processed dataset openly available	81.25	6.25	9.38	3.13
Q57: Raw Data: Share raw data unless constrained by privacy or ethics	75.00	14.06	7.81	3.13
Q58: Ethical Considerations: Address ethical concerns related to data sharing	82.81	3.13	7.81	6.25
Q59: Access and Licensing: Specify access and licensing terms for shared data	89.06	4.69	3.13	3.13