On the Computational Reproducibility of Human-Computer Interaction

Olga Iarygina
olia@di.ku.dk
Department of Computer Science,
University of Copenhagen
Copenhagen, Denmark

Kasper Hornbæk
kash@di.ku.dk
Department of Computer Science,
University of Copenhagen
Copenhagen, Denmark

Aske Mottelson
asmo@itu.dk
IT University of Copenhagen
Copenhagen, Denmark

Abstract

An increasing number of HCI researchers have embraced open science ideas, such as sharing data and analysis code. However, such practices are only meaningful when the shared data and code enable other researchers to reproduce and reuse the reported findings. To investigate the reproducibility of HCI research, we identified all CHI papers that have shared study data and analysis code, and attempted to reproduce the results. We were able to fully reproduce 49% of the papers. We surveyed and interviewed authors, asking them to assess the reproducibility of their own work and to reflect on their motivations and obstacles in doing open science. We discuss what improves and hinders reproducibility and provide recommendations on how to increase reproducibility rates in HCI. While the value of replicability remains contested in HCI, we argue that the more modest goal of reproducibility is desirable.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; • **General and reference** → *Empirical studies*; • **Applied computing** → Digital libraries and archives.

Keywords

Reproducibility, open science, metascience

ACM Reference Format:

Olga Iarygina, Kasper Hornbæk, and Aske Mottelson. 2026. On the Computational Reproducibility of Human-Computer Interaction. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3772318.3791129

1 Introduction

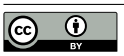
Scientific progress depends on the ability of researchers to verify and build upon each other's work [9]. Open science provides a framework for making research processes and outputs accessible, enabling others to examine, reuse, and extend existing materials in ways that cumulatively advance knowledge [34, 36, 37].

Open science practices are gaining traction, driven in large part by the replication crisis, which has increased the demand for transparency in research [2]. Researchers are increasingly encouraged to preregister studies and to share data and analysis code. The HCI community has partly embraced open science ideas, for example, through continuous advocacy for increased transparency in HCI research [3, 55, 61]. While open science practices are present in HCI, they remain voluntary and far less common compared to other scientific fields [14]. Consequently, discussions of scientific openness in HCI research, or the lack thereof, persist. There remains substantial potential in HCI to improve open science practices [3, 6]. The scholarly discussions of openness in HCI research expose a lack of replications [7, 22], insufficient reporting [26, 54–56], and methodological imprecision [23, 31, 53]. Criticism of open science practices in HCI extends beyond quantitative research, as significant variation in openness is also evident within qualitative studies [51].

'Replicability' and 'reproducibility' are two central ideas in open science. 'Replicability' refers to whether a new study, independent of the original, arrives at the same conclusions by using similar methods [5]. Replicability has been the source of several discussions in HCI [e.g., 7, 22] and has prompted numerous calls to action [59–62]. Replicability is an essential element of scientific progress, but may face challenges in HCI due to increasing familiarity of users with technologies, technological advancements [11], and many other factors that may limit the comparability of studies over time.

'Reproducibility' refers to whether scientists can produce the exact same results of a study, if provided with the study's original data and analysis code [5]. In computational research disciplines, reproducibility is often concretized as *computational reproducibility*, meaning that reproducibility is achieved by rerunning computer code used to generate research findings [50]. Reproducibility is the minimum level of credibility we should expect of published findings [20]. Despite being a more modest aspiration than replication, it has received far less attention in HCI. The HCI field, and in particular experimental HCI research, appears particularly suitable for an increased commitment to reproducibility. The field's computerized experiments, automated data collection, a practice of sharing open-source prototype code to enable reuse, and a widespread use of programming to analyze data and generate figures provide a strong foundation for embracing reproducibility.

To gauge the state-of-the-art of reproducibility in HCI research, we manually attempted to computationally reproduce the 76 CHI papers that over the past 17 years shared data and analysis code. We selected CHI as a representation of high quality HCI research, as the conference is commonly considered the most prominent in the field, and at the same time represents a wide variety of scientific views, methods, and areas of interests. Following our reproduction attempts, we conducted a survey of authors of the papers with openly shared research data to understand their motivations and hindrances in sharing data and analysis code. From the



This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '26, Barcelona, Spain
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791129

reproduction attempts and the survey, we identified particularly well-maintained and reproducible HCI research repositories. We conducted semi-structured interviews with the researchers who created these repositories and asked them to share their reflections on how to increase reproducibility in HCI.

The results from the reproduction attempts, the survey, and the interviews suggest that the community lacks clear guidelines and stronger incentives to encourage authors to make their work reproducible. Reproducibility does not appear to be universally valued, and researchers sometimes face resistance from collaborators when pursuing open science. Authors highlighted areas for improvement, including clearer instructions on how to use repositories, better documentation, ensuring data availability and consistent naming, and asking others to execute their code.

In summary, approximately one in a hundred CHI papers openly share materials allowing reproduction attempt. Approximately half of the openly shared research materials offered direct reproducibility (i.e., we were able to compute the same numerical results as presented in the papers), revealing a concerning quality of openly shared materials in HCI.

2 Background

Open science is increasingly recognized as a cornerstone for building an efficient, progressive, self-correcting, and credible science [24, 33]. Researchers have begun to explore ways to make scientific practices more open, transparent, and credible, building on the crisis of confidence in science, and recognizing that many reported findings cannot be replicated. For example, in 2017, *Transparency and Openness Promotion Guidelines* were developed [19]. The guidelines list 'computational reproducibility' as a recommended verification practice for empirical research claims. An increasing number of journals now encourage that research data and analysis are shared [35].

For scientific research to be computationally reproducible, both data and code must be shared, accessible, and understandable [40]. To enable this, papers must indicate where the code and data can be found; the data should be interpretable (e.g., labeled or accompanied by a codebook), and the code should be annotated. Finally, the code needs to be executable, so that "the independent party can check that they obtain the same results using data and code deposited in a trusted repository" [19].

In recent years, many studies have investigated how often research data and code are shared and whether reported results can be reproduced. Hardwicke et al. [20] evaluated reproducibility of articles published in the journal *Cognition*, where data sharing is mandated, and could reproduce 31% of the articles without the assistance of the original authors. Hardwicke et al. [20] experienced that for some papers it took upwards 25 hours to reproduce the results. Stodden et al. [50] retrieved the data for papers published in *Science* between 2011 and 2012, and were able to reproduce the conclusions for 26% of those. Finally, Obels et al. [40] analyzed the reproducibility of registered reports in psychology, a publishing format in which the methods and analyses plans are reviewed and accepted prior to data collection. In registered reports, sharing data and code is required. The authors were able to reproduce 58% of the reported findings. Limited reproducibility rates were also found in political [47] and geoscientific research [30].

In HCI, there is growing recognition of the values of openness, transparency, replicability, and reproducibility. The submission guideline at CHI 2026 states that *"Where relevant, authors are strongly encouraged to provide supplementary materials to support practices around research reproducibility as much as possible"*[1]. ACM has also introduced an "Artifact Review and Badging System" aimed at encouraging researchers to share research materials. The RepliCHI workshops and SIG panels held between 2012 and 2014 focused on questions about the necessity of replications [59–62]. Cockburn et al. [8] brought the field's attention to preregistrations– a practice of documenting the hypotheses and planned analyses of a study prior to data collection. Preregistration is based on the expectation that data and analyses will be shared after the study's completion. Several articles have called for more transparent and reproducible reporting, with a particular focus on statistical analysis [8, 12, 15, 16, 26, 56].

Despite these efforts, shared data and analysis scripts, preregistrations, and replications remain rare in HCI. For example, Echtler and Häußler [14] found that among all CHI papers published between 2016 and 2017, only 5% included publicly available prototype source code. Abbott et al. [1] showed that, in the health and wellness subfield of HCI, research data are rarely published or are only shared upon request. Hornbæk et al. [22] found that purposeful replications are nearly absent in HCI. Wacharamanotham et al. [55] found that sharing research materials and data is uncommon, partly due to misunderstandings about its purpose. These low rates of sharing have been criticized for hindering replication efforts [1, 22]. In addition, at present, none of HCI venues have officially adopted the research transparency guidelines [19].

LeVeque [32] and Barnes [4] explored the concerns that prevent researchers from sharing source code. They found that researchers are primarily concerned about intellectual property, the effort required to make code publishable, and the fear that code may not be runnable in the future. Additionally, there is a lack of understanding of the value of sharing materials, and the practice is therefore still uncommon. Wacharamanotham et al. [55] investigated concerns specific to sharing data in HCI. The authors observed that some researchers refrain from sharing data, believing that publishing at CHI is already challenging enough and that sharing data might further diminish their chances of acceptance. Other authors highlighted the effort required to prepare shareable materials. It was also mentioned that the purpose of sharing is unclear when there is no demand for the data.

In recent years, it has been suggested that HCI should not only promote and endorse transparency [12, 13, 56], but also contribute to it empirically. However, to our knowledge, no deliberate replications or systematic reanalyses of existing data have been conducted. The exception is Dragicevic et al. [13], who reanalyzed five CHI papers for which data and code were available and augmented their analyses with interactive widgets to demonstrate a tool for transparent statistical reporting.

While the topics of replication and reproducibility receive considerable attention, we do not know whether these discussions are reflected in practice, or whether sharing data and code actually facilitates reproducibility.

[1] https://chi2026.acm.org/authors/papers/

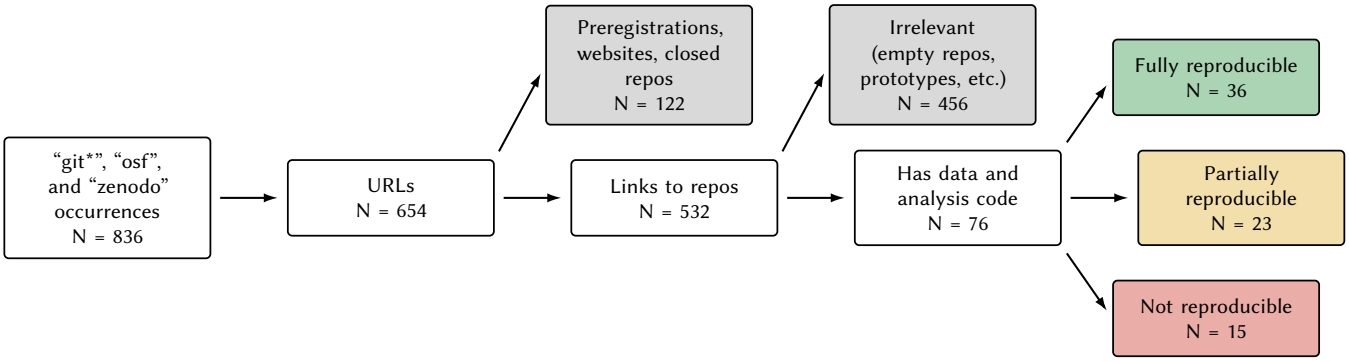


Figure 1: Flowchart summarizing the processing of open science repository occurrences, indicating how many led to accessible repositories and the extent to which they were reproducible.

3 Reproducing CHI

We systematically attempted to computationally reproduce all CHI papers with open repositories that included data and analysis code. Our aim was to evaluate the extent to which computationally generated results at CHI are reproducible and to identify obstacles that can help highlight opportunities to improve reproducibility in HCI.

3.1 Inclusion criteria

Wacharamanotham et al. [55] found that Github is the most popular platform for sharing research materials among CHI researchers. GitHub is usually not recommended for storing research materials as Github repositories can be changed after publication and lack a long-term persistence plan [55]. As an alternative, the authors proposed to use OSF or Zenodo which are compliant with FAIR principles (findable, accessible, interoperable, and reusable) [57]. In the present paper we therefore decided to focus our analysis on repositories hosted on GitHub, OSF, and Zenodo, as these three platforms together appear the most frequently used among CHI researchers. GitHub was launched in 2007, followed by the introduction of Zenodo and OSF in 2013. Consequently, we limited our search to the years 2007–2024 to ensure coverage of all three platforms from their respective inceptions.

To identify repositories that contain open data and analysis code, we extracted all GitHub, OSF, and Zenodo links from the full text of all CHI papers published from 2007 to 2024. We did not analyze supplementary materials hosted in the ACM Digital Library because they seldom contain code, and are not compliant with the findability principle, which requires research materials to be searchable and accessible via a URL referenced in the paper [55].

To extract relevant repository links, we first attempted to collect all URLs from the papers with a script. However, this approach led to false negatives, as repositories were sometimes inserted into the text as embedded hyperlinks. To address this, we searched for all occurrences of "git*," "osf," or "zenodo" in the full texts of papers using a script. This yielded 836 instances, where an instance was defined as a textual reference to one of these platforms, either as plain text, a URL, or a hyperlink. From these, we extracted links where available, resulting in 654 unique URLs. The first author manually coded each link (see OSF[2]). The links included preregistration materials, open-source tools, supplementary materials, project websites, package documentation, and other information. We then filtered the coded URLs to retain only those repositories containing both publicly available data and analysis code related to the findings reported in the corresponding paper. This process yielded 78 links (54 to OSF and 24 to GitHub). Two pairs of links referred to the same datasets hosted on different platforms, and we therefore removed two duplicates. The final dataset consisted of 76 unique repositories. An overview of extraction and filtering process is shown in Figure 1.

3.2 Analysis

We attempted to reproduce the findings of 76 papers with repositories that met our inclusion criteria. We followed the workflow outlined in Figure 2. For each repository, we examined documentation, data availability, code executability, and whether the reported results could be reproduced (see Table 1).

In some cases, error messages occurred when executing scripts from the repositories. We attempted to resolve these errors, but if we spent more than 45 minutes troubleshooting without making progress, we considered the reproduction attempt unsuccessful, as per Obels et al. [40]. In addition, we compiled qualitative notes for each repository. We documented the technical difficulties encountered with each repository, any modifications made to the code, and any discrepancies between the paper and the code output. If the reproduction attempt failed, we recorded the likely cause of failure. In addition, we documented noteworthy repository practices that proved especially helpful for reproduction.

We considered a paper fully reproducible if we were able to obtain exactly the same results as reported in the article, with no discrepancies in our computer generated outputs and the reported outputs. We assigned this code even when we had to resolve errors in the analysis scripts, update folder paths, or infer variable and label correspondences in the absence of a codebook. If the output of scripts were statistical results, we compared numerical outputs generated by the analysis code and those reported in the paper. If the output of scripts were figures, we examined whether the

[2] https://osf.io/uhq8s

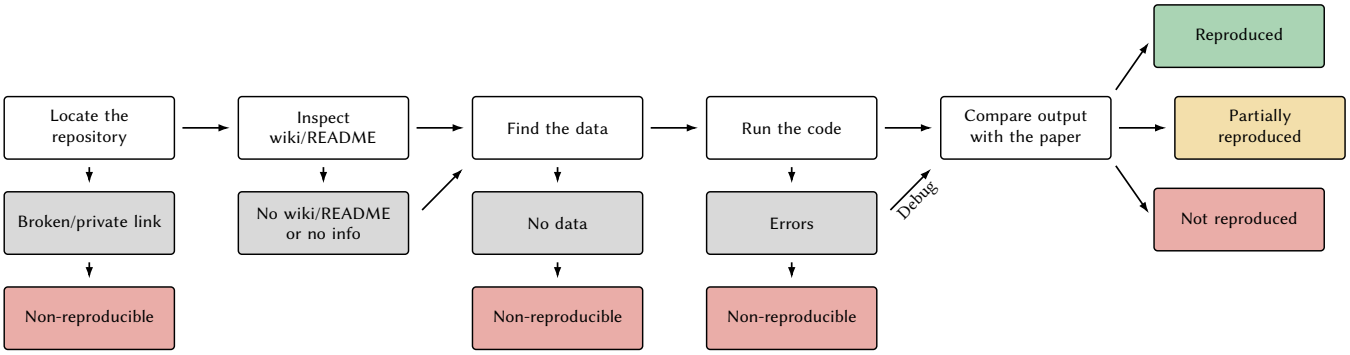


Figure 2: Flowchart summarizing the step-by-step process of reproducing a repository, indicating pitfalls which led to failed reproducibility attempts.

reproduced figures matched those in the publication. We coded a paper as partially reproducible if only part of the analysis code could be successfully reproduced, and in such cases noted the extent to which the results were reproduced. Finally, we coded a paper as non-reproducible if the analysis scripts could not be executed despite troubleshooting, or if none of the results matched those reported in the article. Our coding is available in OSF[3].

3.3 Results

Documentation files. Among the 76 repositories, 55 included a README or Wiki describing the repository. Four of these contained only the paper abstract or a simple "data and code" statement. Twenty-six files with documentation only described the repository structure and its files. Fourteen README files provided instructions on how to use the repository or the order in which to execute the scripts. Among the remaining repositories, three detailed updates made after the paper's publication, three specified software requirements, two contained a codebook for the data, two provided additional resources (e.g., links to an *Observable notebook*), two explained how the scripts related to sections of the paper, and one described deviations from the preregistration. One repository included an explicit note about limited reproducibility due to protected data, although the available data were sufficient to reproduce the main findings.

Data. Almost all repositories (71 out of 76) contained the data relevant for the analysis. However, 13 of these had incomplete data, meaning that some data files needed for the analysis were missing or that data lacked crucial data points (e.g., columns). Among all repositories, only 25 included a codebook detailing the meaning of the data. We examined the data files in all repositories and coded whether they were understandable with or without a codebook. This included cases where variable names were sufficiently clear or where the data were easily interpretable in relation to the paper (e.g., x, y, z coordinates in a movement-tracking dataset). We coded files from 55 repositories as understandable. Most of the data were provided in mainstream formats, such as `.csv`, `.xlsx`, or `.json`. A few repositories contained less widespread formats in data analysis, such as `.rds` (RStudio), `.sav` (SPSS), `.pkl` (Python), or `.dsv` (Microsoft SQL server).

Analysis code. The programming languages used for data analysis were almost exclusively Python and R. In addition, we found two scripts written in SPSS, one in MATLAB, and one in JASP. Not all of the scripts could be executed. We were able to fully run the data analysis scripts from 44 repositories (58%), while 20 scripts could be executed only partially (26%). The reasons for failure varied. In some cases, a required data file was missing from the repository, or specific data columns needed for the analysis were absent. In other cases, custom-defined functions returned errors, and multiple attempts to troubleshoot the code were unsuccessful, particularly when the function was undocumented. In a few instances, the scripts relied on packages and libraries specific to a particular operating system (e.g., the use of Python package `win32com`, which can only be run on Windows).

Reproducibility. We tried the reproduction on a MacBook Pro, with Apple M3 processor, with 16 GB RAM, running macOS Sequoia 15.4.1. We fully reproduced the results of 36 papers, which represents 49% of the sample. We marked 23% of repositories as partially reproducible, with varying levels of success: most of the results were reproduced (12 papers), about half of results (5 papers), and very few results (6 papers). Twelve papers could not be reproduced at all. As mentioned earlier, some papers were not reproducible because of described, yet, missing data files (or parts of data files) required to execute the scripts. This was the case for nine of the non-reproduced papers. In other cases, the code produced persistent errors, despite repeated troubleshooting attempts. For partially reproducible papers, the main reasons were missing data for part of the analysis, missing code for some central findings, errors in sections of the code, or, most commonly, discrepancies between the code output and the results reported in the paper.

3.4 Reflections

Next, we summarize our notes on what made repositories difficult to use and which practices were especially helpful. These reflections are based on notes from each reproduction attempt.

[3] https://osf.io/uhq8s

Complications occurred repeatedly. Many repositories included multiple scripts without indicating the execution order, forcing us to reconstruct both the sequence and the link to paper sections. Fixing path structures was the most frequent correction and became cumbersome when helper scripts also contained paths. The lack of comments in the code made debugging harder, especially when custom functions were undocumented. Large outputs (e.g., factor analyses) further complicated locating reported values. Overall, the greatest challenge was linking code to the paper. In some cases, locating a single reported value in uncommented script took up to half an hour.

We also encountered practices that facilitated reproduction. Detailed READMEs describing folder structure, software, and script order were particularly helpful. RStudio project files (`.Rproj`) simplified path management. Workflows were easier to follow when explicit links between code and paper appeared in the README, file names, or script comments. Executed Python Notebooks and RMarkdown files allowed direct comparison of outputs and error tracing. Some authors even shared the paper in markdown with embedded code chunks. Finally, specifying software requirements, session information, and package versions enabled recreation of the original environment and substantially improved reproducibility.

Dimension	Successful outcome	Pitfalls
Locate repository	Provided and working link to the repository.	Broken link, private or empty repository.
Documentation	Presence of a README or Wiki and its contents (e.g., structure of the repository).	No README or Wiki, or content is not related to the repository (e.g., paper abstract only).
Data availability	The relevant data are in the repository.	Data are claimed but missing or incomplete.
Data format	Data are in accessible formats (e.g., `.csv`, `.xlsx`, `.json`, `.txt`).	Software-specific data formats (e.g., `.rds`, `.sav`, `.pkl`, `.dsv`).
Codebook	Variables are explained in a codebook.	No codebook.
Data clarity	Codebook or variable names are commented or self-explanatory.	No codebook and data are not explained.
Programming language	Language(s) used for analysis (e.g., Python, R, Matlab).	Analysis code is claimed but missing.
Analysis executability	Analysis code runs end-to-end without modifications or with minor fixes.	Missing data, persistent errors, OS-specific packages, no order of scripts execution.
Reproducibility	Outputs from the code match results and figures in the paper.	Missing data/analysis code, non-executable analysis, discrepancies with reported results.
Reproduction time	Time required to reproduce the analysis.	Excessive time spent troubleshooting due to missing documentation or errors.

Table 1: The dimensions we used to analyze repositories.

4 Survey and interview

Our attempts to reproduce CHI showed that only half of the repositories with open data and analysis code were fully reproducible. However, our coding of the reasons does not capture the authors' own views on the obstacles and the likely causes of failed reproduction, nor their motivations for sharing materials. To connect repositories with authors' experiences, we surveyed the authors of the 76 identified repositories. The survey explored how researchers organize and share materials, how they perceive the reproducibility of their work, and what motivations and challenges they experience. In addition, we invited authors of repositories we subjectively marked as exemplary to participate in semi-structured interviews, to gain deeper insights into workflows for preparing reproducible data and analysis code that could inform other researchers. We deliberately focused on cases of successful practice to examine the concrete workflows, tools, and organizational strategies these authors used to create and maintain highly reproducible repositories. By studying these positive examples, we aimed to distill actionable practices that could inform and support the broader research community in developing more reproducible data and analyses.

4.1 Survey design

We conducted an online survey, where we invited all authors of the papers that we attempted to reproduce. The survey contained five sections. In the first section, we informed participants that we had attempted to reproduce the results of their paper but did not disclose the outcome of the reproducibility attempt. We asked whether they believed another researcher could reproduce the results of their paper. If they thought that we were unable to reproduce findings, we asked them to explain why.

In the second section, we asked about sharing practices and how responsibilities for data sharing are distributed within research teams. The third section focused on motivations for sharing data. Section four asked whether authors had ever attempted to reproduce someone else's analysis, as we speculated such experiences could influence how they make their own data available. The final section contained demographic questions, including seniority level and years of research experience. The full list of survey questions is provided on OSF[4].

We emailed all the authors of the papers included in the sample. To collect the email addresses, we scraped the papers using the `pdftools` package in *R*. If the authors did not respond within 10 days, we sent a reminder. If we received an auto-reply that the email was not active anymore, we manually searched for an up-to-date email. We conducted the survey using Qualtrics. We used individual links to dynamically insert references to the author's paper and repository into the questionnaires, so that if an author had multiple

[4] https://osf.io/uhq8s

papers with open data published at CHI, they knew which paper we were surveying about. This setup allowed us to track which authors responded, but we did not link individual responses to specific authors[5]. The survey took approximately 10 minutes.

We received 63 complete responses out of 305 invitations. Respondents had a mean of 10.7 years of research experience (SD = 7.8). 49 respondents held a doctoral degree, 11 a Master's degree, two a Bachelor's degree, and one reported other educational background. In some cases, multiple authors from the same paper responded, while in others, authors declined participation because they felt they were not sufficiently involved in the data collection or analysis to answer our questions. Several authors who felt unqualified forwarded our invitation to collaborators who they considered more appropriate respondents. In total, we received responses corresponding to 45 out of 76 papers, covering 59% of the sample.

4.2 Interview design

In the interviews, the same overall questions guided the conversation, but the format allowed digressions based on unplanned but interesting points. Additionally, we asked authors about their workflows for preparing data and analysis code, and for organizing the repository. The full list of interview questions is provided on OSF.

We invited the first authors of seven papers whose repositories we had subjectively judged to be exemplary based on the reproduction attempt. We conducted interviews with four authors who responded to our invitation. The interviews lasted approximately 45 minutes.

4.3 Data analysis

We analyzed only complete survey responses. Additionally, one author informed us by email that they had only been exploring the survey questions and had entered placeholder answers. When they reached the last page, they were unable to return and complete the survey properly. We discarded this response and sent the author a replacement link. For quantitative questions, we calculated proportions (the number of responses in a given question divided by the total number of responses). For open-text questions and interview transcripts, all authors of this paper collaboratively grouped the responses into descriptive categories. In the results, we present both representative and exceptional quotes based on these categories.

We present the survey and interview data together. In the following, quotes from survey participants are marked with (S#) and quotes from interview participants with (I#), where the number refers to the anonymized participant ID (e.g., S10 or I4).

4.4 Results

Figure 3 shows an overview of the survey questions. Table 2 presents a summary of the free text survey responses and the interview.

Perceived reproducibility

Most of the respondents (60%) expected that we could fully reproduce the results of their paper using the materials they provided. Fewer (33%) expected that we were able to reproduce only the key results. Some authors (6%) thought that we would not be able to reproduce their results at all.

If the authors thought that we could not fully reproduce the results of their paper using the materials provided, we asked them why. Four authors explained that some material was not shared because of privacy concerns, and seven participants did not share all the data needed for reproducing results, for example, unprocessed tracking data. Beyond privacy, one author explained that they did not share the data because they were *"concerned (raised by a participant) that our quality labels may be perceived as offensive or stigmatizing"* (S60). Two authors pointed out possible difficulties related to using shared files:

> *While we shared all of the materials, if a researcher wanted to re-run the study they would face some challenges – specifically as we used Qualtrics (which is a proprietary survey platform) the qsf file may not be usable by everyone.* (S1)

Five authors explained that they did not test the shared materials, and therefore, they remained uncertain about their reproducibility. Two participants admitted that at the time of publishing a paper

Theme	Key results
Perceived reproducibility	60% expected full reproduction, 34% expected partial, and 6% expected that none of the results will be reproduced. Perceived reproducibility partially mismatched our reproduction attempt.
Reasons for not sharing	20% did not share some materials. The reasons are: privacy, sensitive annotations, lack of skills, lack of time, no understanding of value of sharing.
Responsibilities	In 42% of cases, data and analysis were prepared jointly. Authors are often responsible for their own part; some teams check each other's code or test analysis on multiple devices.
Reproducibility experience	65% had tried reproducing others' work. Most reported no effect on their own workflow, but some said it influenced their practices (e.g., importance of documentation, serving as inspiration).
Motivation to do open science	Belief that openness improves science, credibility, and impact. Some noted fears (exposure of mistakes, vulnerability) but valued accountability. Replication crisis in psychology cited as cautionary tale.
Lack of incentives	Reported lack of standards, guidelines, infrastructure, and recognition. Suggested solutions: guidelines, independent reproducibility checks, online execution environments, and recognition from CHI.

Table 2: Overview of survey themes and main findings from authors of reproduced papers.

[5]The details of the process are described in the Qualtrics documentation, under the "Anonymized responses" section: https://www.qualtrics.com/support/survey-platform/survey-module/survey-options/survey-protection/

they did not have skills for data sharing and version control. Among the rest, one person was not responsible for the analysis, one person was unsure that every detail from the paper would be present in the repository, and two authors provided irrelevant answers.

We were interested in how much actual reproducibility matched perceived reproducibility, and connected survey data with our reproducibility attempt (Figure 4). Among 38 authors who thought that their paper was fully reproducible, we could fully reproduce 22. Surprisingly, we manged to fully reproduce 16 papers from the 21 authors who indicated that their paper was only partially reproducible; and we were able to at least partially reproduce half of the papers that authors believed were not reproducible at all. Analysis of free texts suggested that authors gave this response because they either thought that we were reproducing qualitative findings as well; because they were unsure as the repository was released long ago; or because they were not responsible for the analysis.

Reasons for not sharing

Nineteen percent of participants reported being unable to share some materials relevant for the study in the paper. In seven cases the reason was privacy concerns. For example, one author explained:

> *We had raw physiological data (heart rate, HR variability, skin conductance, etc.) If I recall correctly, we did not have ethics approval for sharing this, or thought it might be too personal of data to share.* (S46)

One author said sharing is difficult and resource-consuming, and three authors admitted that they did not know why they did not share some data. One author specified that after participation in our survey they are working on uploading missing data.

If authors answered that they not always share their materials (37%), we asked them how they decide when to share, and when not. Out of 23, nine respondents always share data when there are no privacy concerns. Three people said that they would like to share everything but they do so only when time allows. Two people responded that they share only when they see value in it, and *"when I think I can claim the tool or analysis method scripts as a contribution"* (S23). Peculiarly, two respondents complained about a sharing experience at CHI:

> *My reviewers were incapable of commenting on statistical details. I doubt they or similarly experienced colleagues would have the statistical expertise to make out which result from the reproduction corresponds to the paper results.* (S45)

Finally, seven people reported that they share only when they are the leading authors, because sometimes collaborators *"push back on sharing"* (I3).

> *I would rather continue collaborating than stop collaborating over such a difference in opinion.* (S6)

Responsibilities

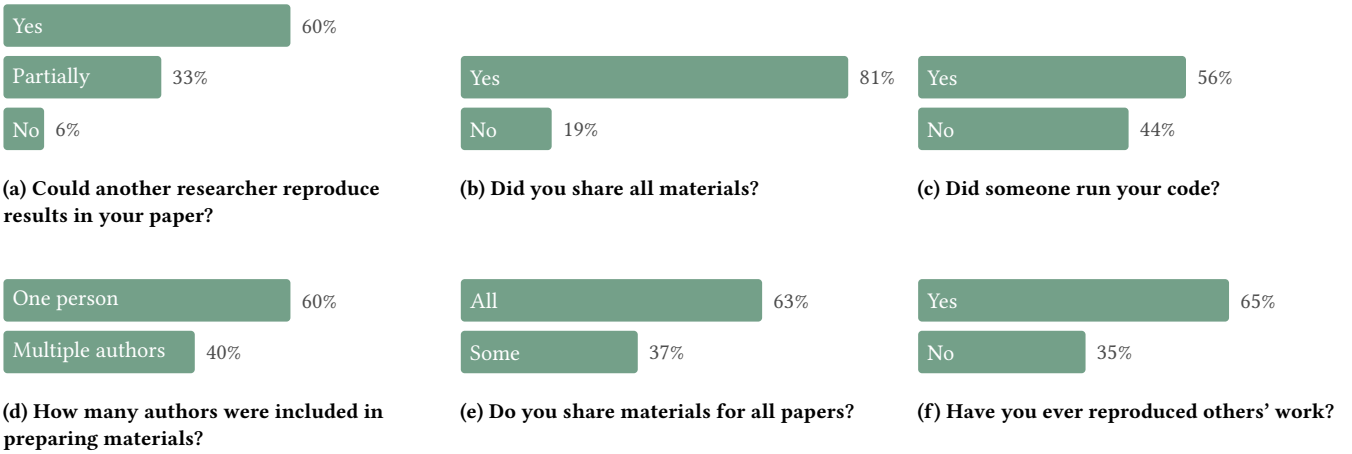
In 40% of responses, materials were prepared by multiple authors. We were interested in how the responsibilities were distributed within the team. Fifteen authors said that each author is responsible for his own part of work. Eleven authors described that they have a practice of checking each other's code before publishing.

> *Another author and me analysed different components of the data and then I tried running all scripts and replicating all results and polished the scripts for reproducibility. Then a third author tried running all of the code and gave feedback on the README scripts. It was very much a joint effort.* (S2)

A few authors try to ensure that the script *"works on different devices (e.g., using different instances of the same statistics software to see if all packages run as expected)"* (S50).

Reproducibility experience

We assumed that if authors ever tried to reproduce someone's work, this experience could influence their own practices in sharing materials. Out of 65% of authors who had reproducing experience, many authors said that it did not have any influence on their own practice. However, eight researchers stated that open science repositories and practices served as an example for them.



(a) Could another researcher reproduce results in your paper?

(b) Did you share all materials?

(c) Did someone run your code?

(d) How many authors were included in preparing materials?

(e) Do you share materials for all papers?

(f) Have you ever reproduced others' work?

Figure 3: Survey responses on reproducibility and open science practices among CHI authors ($N = 63$).

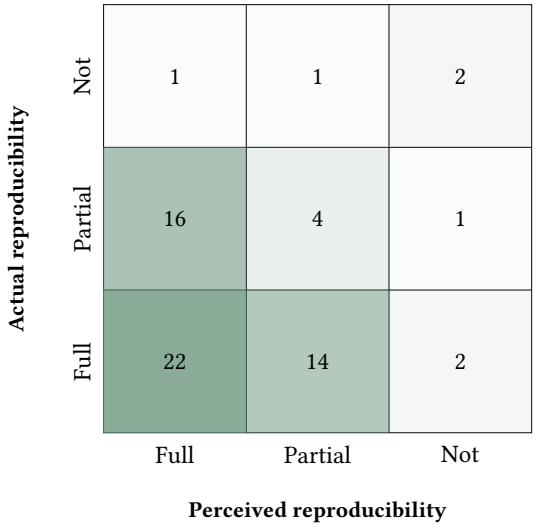


Figure 4: Relation between perceived and actual reproducibility of papers, with numbers of papers in each cell ($N = 63$).

> *It was one of the first analyses I reproduced to see how the work was done and to gain some insights into how I can share my own research from a reproducibility point of view. It was a really well-documented analysis and it gave me several ideas on how to do the same in my own work.* (S32)

Twelve researchers reported that trying to reproduce others' work made them understand *"how important proper documentation is and how detailed you have to be"* (I1), because they *"don't want others to feel so painful to reproduce my work"* (S60).

Motivation

We asked authors what motivates them to follow open science practices. Many authors stated that they *"believe it is a better way to do science"* (S21), and *"this is a crucial element of research"* (S7). Others believe in the importance of transparency and reproducibility. Some authors explained that open data and analysis increase the credibility of the research, and they feel *"responsibility to provide proof to their claims"* (S33).

> *I lost trust for many results and claims of some previous studies. So I think sharing data and code is the only way to allow other researchers to test claims in the future, but it's also a way to increase our research responsibility – as we cannot make claims that are not supported by public data.* (S56)

Two authors described that it is *"scary"* (I4) and *"vulnerable"* (I2) to share analysis because they could *"made a mistake and then other people might find it"* (I4). However, they still share as if analysis is not shared, the mistake might be never found out, and they would *"rather have people point out if something is not working"* (I4). One author also pointed out that even if the analysis is correct, interpretations could be different:

> *Even in controlled experiments you might have different interpretations because, you know, statistics. I shared the code. People might say, okay, I would not go like this, I would go like that. And I like that. I like the fact that people might confront you, but it is always based on data.* (I2)

Another three authors said that open materials make their work more impactful, provide *"an additional or stronger contribution"* (S23), and they *"want others to be able to build on my work"* (S50).

> *I think like sharing the process from raw data to what we actually use for analysis in the paper is super important. Because otherwise you miss a key step if you can't turn the raw data into what we used in the paper. You can't reproduce it either. The raw data frame by itself isn't my intellectual contribution. It's all the metrics I derive from it.* (I1)

Reflecting on the replication crisis in psychology, one author stated that she does open science because we should learn from replication crisis in HCI:

> *We just cannot have that. It's a giant waste of taxpayer money if we publish bullshit to play the game of staying hired in academia instead of publishing to advance our understanding of how the world works.* (S58)

Incentives

Finally, we asked researchers if they lack any incentives from conferences, journals, and institutions that would foster reproducibility. Some shared their opinion that *"a requirement (to make research reproducible) is the only tactic that I think would have serious effects on whether or not people share data and code"* (S21). Some authors mentioned the lack of resources, such as too little time for preparing supplementary materials at CHI, or lack of independent reproducibility checks:

> *An independent team from the publication outlet or reviewers checking for reproducibility would be awesome – getting some feedback is truly helpful especially as a junior researcher.* (S41)

Researchers repeatedly expressed *"a fundamental lack of standards with sharing and reproducibility"* (S40) and a need for *"a checklist or set of guidelines for best practices to make research reproducible"* (S46). This lack of guidelines was underlined as especially relevant for junior researchers, as there is no specific, universal and easily-findable set of resources:

> *I must say that it took some effort to make our data reproducible and produce a good script. Maybe if there were documentation of everything that is needed to create a well reproducible script it would have made the process a bit faster. I remember searching for templates or documents that had information on what was needed and OSF had some resources that were great to have. But it was a bit difficult to know what to search for in the beginning.* (I4)

Researchers also report a lack of proper infrastructure for sharing materials. They describe OSF as "frustrating to use" (S8) and note

that "OSF could learn much from the GitHub community" (S39). At present, sharing of materials is unstandardized and decentralized:

> *I need a centralised thing that everyone uses. Like GitHub makes it easy for tools. But for data and analysis it feels like there are so many random different services to learn and then forget about.* (S23)

Some researchers would like to have an online way of running code *"(e.g., some kind of Jupyter notebook), because you don't have to download or install anything"* (S7).

Finally, researchers lack recognition for the amount of work needed to do open science. This concerns, for example, the way of assigning conference rewards:

> *As junior level researchers, whether our paper gets a "best paper" or "honorable mention" is often times a matter of pure chance and luck, rather than the quality of the paper itself. It would be great if there could be some publicly acknowledged way, such that people who spent more work on making things transparent and reproducible are acknowledged.* (S52)

While *"not everyone loves badges and says, oh, they're like a token things"* (I1), researchers also *"think it could be useful to propose badges for different levels of transparency and reproducibility"* (S25). One author highlighted that even though CHI 2025 had a checkmark for badges during submission, *"I was searching or looking at proceedings and none of these badges are applied"* (I3).

5 Discussion

We attempted to reproduce CHI papers with open repositories containing data and analysis code. Out of 76 papers, we could fully reproduce 37 (49%). For comparison, the journal *Cognition* reported a reproducibility rate of 31% [20], and an analysis in *Science* reported a reproducibility rate of 26% [50]. Registered reports in psychology, where expectations for reproducibility are especially high, were reproducible in 58% of cases [40]. Thus, CHI does not appear to perform worse than other fields. Yet, irrespective of other fields' low reproducibility rates, what does it mean that only half of CHI papers that openly share data and analysis code were fully reproducible?

In the following sections, we (1) outline why we believe reproducibility is not just important but essential, and discuss what the obtained reproducibility rates imply for CHI; (2) discuss benefits of reproducibility; (3) provide practical recommendations on how to create reproducible repositories; (4) discuss how the CHI community can increase reproducibility rates, and (5) outline limitations of our study.

5.1 Why reproducibility is essential

For many, the ultimate standard by which empirical scientific claims are judged is replication – independent attempts to repeat a study and reach the same conclusions as in the original work [41]. While replication is a commendable effort, it requires conducting an entire study, which is time- and resource-intensive, and therefore not always feasible. CHI is routinely called out for the lack of replication work [22, 59–62]. At the same time, human-computer Interaction as a field, and the CHI conference in particular, strives towards novelty. Novelty requirements make people focus on chasing new work, not replicating the existing studies. In addition, many contributions in HCI take the form of prototypes. Prototypes and small-scale user studies are not well suited for replication (e.g., scientific claims in them are not the main contribution). Finally, even with scientific rigor, external factors may complicate replication in HCI. For example, HCI studies are often affected by temporal factors – for example, users adapt to technologies over time, and the technologies themselves evolve rapidly [11].

In contrast, reproducibility – the ability to rerun the reported analyses and obtain the same results using the data and analysis code provided by the authors – is considered the minimum threshold of research credibility [20]. There should be hardly any reasons why a study, if provided data, does not reproduce (beyond sensitive data omitted). What does it mean, then, that only half of CHI papers with available data and code were fully reproducible? This finding raises serious concerns. When studies suffer from missing data, coding errors, or discrepancies between the reported results and those generated by the code, it becomes unclear which results can be trusted and whether the conclusions of the paper are well-founded [38]. An anecdotal yet illustrative example might be an article from *Nature* that was retracted after the authors discovered a mistake in one sign in the code, which changed the entire conclusion [46]. Such issues increase uncertainty regarding the merit of allocating further resources to replications or building on the existing studies [39].

We found the first mention of an open science repository at CHI in a paper from 2015 [27]. In 9,435 papers published in 2007–2024, only 1% of papers have both data and analysis code available. We acknowledge that HCI is a interdisciplinary field with diverse analytical frameworks, which means that many papers follow practices that do not align with open science ideas. Also, computational reproducibility does not apply to all studies. Yet, the aspiration for reproducibility is much more modest than for replication. Many papers in HCI, including qualitative and observational studies, include data-driven figures or report computed metrics that often allow for reproducibility. Nevertheless, our result show that for 99% of CHI studies, it is impossible to verify the validity of computationally derived claims made. In contrast, many general scientific journals require data availability (e.g., *PLOS*[6], *PNAS*[7], or *Science*[8]) regardless of whether the data are quantitative or qualitative, and strongly encourage code submission when the code has been used, making *all* studies open for validation.

The purpose of this paper is not to force open science practices upon a diverse field as a one-size-fits-all solution. We acknowledge that reproducibility is not uniformly feasible or appropriate, as there might be legitimate constraints: privacy concerns, computational demands, proprietary hardware, or restrictive software licenses. But the default assumption should still be that computationally derived evidence should be reproducible unless there is a valid reason it can not be recomputed by third party. If we care about the credibility of empirical research in HCI, reproducibility might be a standard we should be able to agree on.

[6] https://plos.org/open-science-publishing/
[7] https://www.pnas.org/author-center/submitting-your-manuscript
[8] https://www.science.org/content/page/science-journals-editorial-policies

5.2 Beyond reproducibility

Ioannidis [25] emphasized that reproducibility does not guarantee that published results are correct, valid, or of high quality. Instead, it ensures that others have the opportunity to investigate what was done and to verify the credibility of the claims. However, access to analysis code also makes it possible to assess validity. For example, tracing the computational steps taken to generate results, seeing how outliers and missing values were handled, examining which model parameters were applied, enable assessment of how the paper's conclusions were reached and what limitations they may have. Data and analysis code shared at a submission stage also enable reviewers and readers to detect potential errors in the analysis, correct them, and improve the quality of reported results [18, 49].

Reproducibility also offers benefits beyond credibility and validity checks. Prior work showed that sharing reproducible data and analysis code allows researchers to reuse and build on the materials of others, for instance, by reusing figures, calculating new metrics, or understanding the conversion of raw data into data frames for analysis [30, 40, 49, 55]. Several authors in our study shared experiences when they had to spend considerable time reverse engineering figures or metrics from prior work instead of reusing them directly. For example:

> *The paper is older, so I don't blame the authors. I love Jeremy Bailenson's work. But they don't make their datasets available. And there was a particular metric we wanted to calculate, social synchronization. And I was super excited about this. But looking through the papers they give a few lines, description of the transformations they did to their stuff. And they have these synchronization scores, which come magically kind of out of nowhere. And it was, I mean, it was not reproducible. And I did my best from the text to reproduce what I thought they were doing.* (I1)

This example illustrates how the lack of reproducibility can turn promising opportunities for building on and advancing prior work into time-consuming guesswork.

Reproducibility is essential for research synthesis and metascientific work. When research is not reproducible and its outcomes lose credibility, research synthesis is undermined, as it becomes unclear what results any novel analyses should be compared against or whether the results are suitable for inclusion in meta-analyses [20, 48]. It also becomes impossible to reuse materials for comparative purposes. For example, Dragicevic et al. [13] presented explorable multiverse analysis reports for five existing papers and offered alternative ways of analyzing them (e.g., providing a Bayesian reanalysis of a study originally analyzed with frequentist methods). Such reanalysis would not have been possible if the original papers were not reproducible.

Finally, reproducibility might be beneficial for teaching. For example, Yildiz et al. [63] found that reproduction projects foster critical thinking and methodological awareness among students and researchers. The experience of attempting to reproduce published results allowed learners to confront discrepancies between reported findings and actual implementations, and deepened their understanding of research methods and limitations.

5.3 How to make a reproducible repository

We created a set of recommendations on how to make repositories reproducible, in response to repeated requests in our survey and interviews, and because of the lack of comprehensive guidelines currently existing. The recommendations are based on existing literature [17, 37, 40, 49], the FAIR principles [57], and our own reproduction attempts. Below we provide them in condensed version; the full extended guidelines with examples, templates, and links are available in Appendix and externally on OSF[9]. These guidelines can be used either as a step-by-step instruction when creating a new repository or as a checklist for assessing and improving the reproducibility of an existing repository.

Setting up a repository. The first step toward reproducibility is setting up a repository. Repositories that comply with the FAIR principles [57] include OSF (https://osf.io) and Zenodo (https://zenodo.org). Some researchers working with version control prefer to use GitHub (https://github.com). In such cases, a GitHub repository can be archived on Zenodo to make it persistent and immutable. We recommend using OSF for its simplicity, ubiquity, and integration with other tools. It is important that the repository is findable. To ensure this, researchers should attach the repository link to the paper and add metadata to improve discoverability. Several times we encountered references to repositories mentioned in a paper, but they were private, and, hence, inaccessible. Once the paper is published, the repository should be made public.

Adding documentation. A frequent complication we encountered in our reproduction attempts was the inability to understand the contents of a repository and how to execute its scripts. This can be mitigated by providing clear documentation. Typically documentation is done in the form of a README or Wiki file, which is often the first item a visitor sees. Authors in our survey reported that they were sometimes unaware of the need for such files or of what they should contain. Our reproduction attempts showed that useful documentation includes: (1) a link to the paper, (2) an overview of the repository structure, (3) a description of what each file does, (4) the requirements for running the code, (5) the location of the working directory, and (6) instructions on how to use the repository (e.g., the order in which scripts should be executed).

Data. The most common reason for the failure of reproduction attempt in our sample was missing data files. It is important to ensure that all necessary data are anonymized and included. If some files cannot be shared, this should be explained in the documentation. For reproducibility and reusability, data must be understandable. Column names should be meaningful and ideally a codebook should accompany the dataset to explain variable names, their meanings, and values.

Analysis code. In our reproducibility attempts, it was sometimes impossible to execute or troubleshoot the code. To address this, researchers should provide clear instructions on how to run the scripts (e.g., terminal commands, used software, the order of scripts execution). The most common correction we had to make was fixing broken file paths, which can be avoided by using relative and consistent paths. Troubleshooting also became difficult when the

[9] https://osf.io/uhq8s

purpose of the code was unclear; well-annotated scripts are essential for helping others (and the authors themselves at a later stage) understand the analysis. Finally, because packages are regularly updated and functions may become deprecated or behave differently, it is important to specify the package versions used.

Relation of the code to the paper. The hardest part of our reproducibility attempts was linking the code output to the corresponding parts of the paper. This problem can be avoided by making the connection explicit. For example, by explaining in the README which scripts relate to which sections of the paper, or by adding headers and annotations within the code itself. Another helpful strategy is to name scripts and outputs after the relevant sections or figures in the paper (e.g., `figure1.png`). When using RMarkdown or Jupyter notebooks, providing a knitted HTML or PDF of the executed notebook enables others to directly compare the original outputs with their reproduction attempts and makes potential errors easier to identify. If the analysis produces outputs outside of notebooks, including a folder with expected results for reference can serve the same purpose.

Organization of files and verification. We found that meaningful organization of files in a repository makes it much easier for others to understand and reuse. This includes placing files into well-structured folders (e.g., `data`, `analysis`) and giving the files descriptive names (e.g., `survey_data.csv`, `analysis_script.R`). Once the repository is finalized, it is a good practice to download it and re-run the entire analysis from scratch to confirm that everything works as intended, ideally to ask a colleague to do it. If any changes are made after the paper is published, they should be clearly documented in the README or Wiki file. This ensures transparency and allows others to distinguish between the published version and subsequent modifications.

Additional recommendations. Interviews with the authors of the exemplar repositories provided several useful insights into their workflows. They emphasized the importance of planning for data sharing early in the research process and commenting scripts as work progresses. This not only saves time when preparing materials for submission but also helps in remembering what was done and in debugging the code. Authors of the reproducible repositories also reported that they usually ask colleagues to test their code on multiple devices to ensure it runs reliably across environments. Finally, many noted that experience of reproducing others' work taught them what practices are helpful and what obstacles make reproduction frustrating.

5.4 Recommendations for the HCI community

Multiple researchers in our survey and interviews reported that they lack incentives from conferences, journals, and the broader community to prioritize reproducibility. This lack aligns with findings from other sciences, where similar barriers contribute to the publication of non-reproducible research [37]. Below, we outline several possible directions for improvement.

The CHI 2026 guide for preparing a paper already encourages researchers to make their work reproducible: *"Where relevant, authors are strongly encouraged to provide supplementary materials to support practices around research reproducibility as much as possible."* However, the guide does not explain what reproducibility means or how to achieve it. Outlining the value of reproducibility and providing detailed recommendations might encourage researchers to share data and analysis code, and increase reproducibility rates in the long term. In addition, multiple researchers emphasized that preparing high-quality data and analysis code is time-consuming, and extending the timespan for preparing supplementary materials could increase the amount of materials submitted alongside papers.

Several researchers also noted that reviewers should engage more actively with the provided data and analysis code and offer feedback based on reproducibility checks. While we cannot expect reviewers to always carefully verify all supplementary materials, additional resources could be introduced to support such efforts. For example, SIGPLAN has established a shadow committee for artifact evaluation[10], where a team of PhD students checks the materials for reproducibility. Papers that pass this evaluation receive benefits such as additional presentation time, an extra page to describe the artifact, and a special recognition in the proceedings.

Researchers further reported that awards for reproducible materials, or greater transparency around the use of badges [28, 44, 45], would encourage them to make their work reproducible. For instance, at the time of the CHI 2026 submission deadline, no badges had yet been released for CHI 2025 papers.

Finally, several participants stressed that the community itself could be more supportive of reproducibility. Instead of encouragement, some researchers reported experiencing what they described as *"remarkably toxic discussions on open science that really dominated the landscape"* (S40). Researchers find that sometimes doing open science is even demotivating, as *"at CHI, there's a fair share of reviewers who will always ask for more data/code/materials, without actually attempting to reproduce the results"* (S15) and that the feedback often comes in the form of a *"generically worded critique, that seemed a direct reaction to our having shared data and code on OSF and having pre-registered the study"* (S15) rather than in a form of constructive support.

Building a culture of reproducibility does not happen overnight. It requires time, collective effort, and the gradual integration of new norms into everyday research practice. In HCI, the movement toward open science and reproducibility is young, yet there are encouraging signs. More researchers now share data and analysis code, preregister their studies, and adopt transparent practices (see data in OSF[11]). At the same time, it is important to acknowledge that these changes are incremental and uneven, and that what counts as "best practice" continues to evolve. Researchers themselves often emphasize that practices mature gradually, as illustrated by one of the surveyed authors:

> *One thing to keep in mind is that research methods keep evolving and our practices as well. For example, when you asked if I make all my research open and reproducible, I answered yes because I've been doing so for at least 10 years, but for older research I did not yet pay attention to everything I value today, and while I try to make everything public that I can recover, I was*

[10] https://artifact-eval.org
[11] https://osf.io/uhq8s

> *not able to find all data and all scripts for studies I did 15 years ago.* (S25)

We want to stress that our analysis is not intended to assign blame for past shortcomings but rather to highlight challenges and encourage further progress.

5.5 Limitations

Our study has several limitations. In our reproducibility attempt, we analyzed whether we could achieve the results reported in the paper using the data and code provided by the authors. We did not evaluate the correctness of the statistical analysis or the validity of the study. Therefore, our results reflect only the quality of reporting, not the overall quality of the paper.

We attempted to reproduce the papers using only the data and analysis code openly provided in the repositories. We did not contact the authors if something was missing, if we encountered errors during troubleshooting, or if the repository was closed. The assumed aim of shared data and analysis code is that results should be reproducible and reusable even if the authors cannot be reached, in contrast to making files available only upon request [40]. However, had we contacted the authors, our reproducibility rates might have been higher.

It should be noted that our reproducibility attempts could be limited by our coding expertise. While we have substantial experience with data analysis in R and Python (over eight years), the studies in our sample were very diverse, and some partial reproducibility rates may have resulted from our inability to troubleshoot the code. However, ideally, reproducible analysis code should be both executable and self-explanatory.

Participants in our survey were authors who had already shared data and analysis code and who adhere to good research practices, and our survey is subject to self-selection bias [21]. Therefore, we do not know the experiences and obstacles of researchers who did not share analysis code and data, which might also be crucial for understanding the broader landscape. We also do not know the motivations for not sharing among those who are opposed to open science (and our survey suggests that such researchers do exist). The literature suggests that researchers may withhold data and code due to technical and cultural limitations [43]. These limitations include difficulties in managing data and writing clean code, as well as norms discouraging the release of analysis materials to avoid others "scooping" ideas [43]. Other reasons include fear that transparency may allow others to question conclusions [42], reluctance to respond to inquiries about code [10], or even the inability of authors to reproduce their own results [52]. Finally, researchers may simply lack awareness of, or incentives for, practices that would justify the additional effort required to prepare reproducible materials [37].

6 Conclusion

Reproducibility of the outcomes reported in a paper represents the minimum level of research credibility. To understand the landscape of reproducibility in HCI, we attempted to reproduce the results of all CHI papers that shared data and analysis code. We were able to fully reproduce 49% of the papers. The main reasons for limited reproducibility were missing data, errors in the code, and discrepancies between the generated and reported outputs. To better understand the practices leading to such reproducibility rates, we conducted a survey and interviews with the authors of the sampled papers, asking them about their perceived reproducibility, motivation for engaging in open science, and the obstacles authors face in making data and analysis code reproducible. We found that authors often lack training and clear guidelines on how to make their research reproducible, as well as acknowledgment for their work from organizations and the broader community. Sometimes, researchers do not see value in reproducibility. We discuss the value of reproducibility, outline how the community can support reproducible research, and provide practical recommendations for creating reproducible repositories. We hope this paper will help researchers share data and analysis code in ways that enable others to reproduce their results, reuse their materials, and strengthen the credibility of HCI research.

7 Open Science

We share the results of our reproduction attempts, the analysis of the repositories mentions, and the scripts used to generate figures at https://osf.io/uhq8s. Ironically, we do not share certain data files in our open science repository. We do not share free text results of the survey and interview transcripts. The sample of the papers we attempted to reproduce is small, and disclosure of the survey and interview utterances could breach promised anonymity, which was essential for collecting honest experiences from participants.

References

[1] Jacob Abbott, Haley MacLeod, Novia Nurain, Gustave Ekobe, and Sameer Patil. 2019. Local Standards for Anonymization Practices in Health, Wellness, Accessibility, and Aging Research at CHI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300692

[2] Jens B. Asendorpf, Mark Conner, Filip De Fruyt, Jan De Houwer, Jaap J. A. Denissen, Klaus Fiedler, Susann Fiedler, David C. Funder, Reinhold Kliegl, Brian A. Nosek, Marco Perugini, Brent W. Roberts, Manfred Schmitt, Marcel A. G. Van Aken, Hannelore Weber, and Jelte M. Wicherts. 2013. Recommendations for Increasing Replicability in Psychology†. *European Journal of Personality* 27, 2 (2013), 108–119. arXiv:https://doi.org/10.1002/per.1919 doi:10.1002/per.1919

[3] Nick Ballou, Vivek R. Warriar, and Sebastian Deterding. 2021. Are You Open? A Content Analysis of Transparency and Openness Guidelines in HCI Journals. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 176, 10 pages. doi:10.1145/3411764.3445584

[4] Nick Barnes. 2010. Publish your computer code: it is good enough. *Nature* 467, 7317 (2010), 753–753.

[5] Karl Broman, Mine Cetinkaya-Rundel, Amy Nussbaum, Christopher Paciorek, Roger Peng, Daniel Turek, and Hadley Wickham. 2017. *Recommendations to Funding Agencies for Supporting Reproducible Research.* Technical Report. American Statistical Association.

[6] Tatiana Chakravorti, Sanjana Gautam, and Sarah M. Rajtmajer. 2025. Embracing Transparency: A Study of Open Science Practices Among Early Career HCI Researchers. arXiv:2410.04286 [cs.HC] https://arxiv.org/abs/2410.04286

[7] Andy Cockburn, Pierre Dragicevic, Lonni Besançon, and Carl Gutwin. 2020. Threats of a replication crisis in empirical computer science. *Commun. ACM* 63, 8 (2020), 70–79. doi:10.1145/3360311

[8] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3173715

[9] Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349, 6251 (2015), aac4716. arXiv:https://www.science.org/doi/pdf/10.1126/science.aac4716 doi:10.1126/science.aac4716

[10] Christian Collberg and Todd A. Proebsting. 2016. Repeatability in computer systems research. *Commun. ACM* 59, 3 (Feb. 2016), 62–69. doi:10.1145/2812803

[11] Alan Dix. 2020. *Statistics for HCI: Making Sense of Quantitative Data.* Springer Cham, Switzerland. 181 pages. doi:10.1007/978-3-031-02228-9

[12] Pierre Dragicevic. 2016. *Fair Statistical Communication in HCI.* Springer International Publishing, Cham, 291–330. doi:10.1007/978-3-319-26633-6_13

[13] Pierre Dragicevic, Yvonne Jansen, Abhraneel Sarma, Matthew Kay, and Fanny Chevalier. 2019. Increasing the Transparency of Research Papers with Explorable Multiverse Analyses. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3290605.3300295

[14] Florian Echtler and Maximilian Häußler. 2018. Open Source, Open Science, and the Replication Crisis in HCI. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3170427.3188395

[15] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Paweł W. Woźniak, and Albrecht Schmidt. 2019. Gamification in Science: A Study of Requirements in the Context of Reproducible Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300690

[16] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Paweł W. Woźniak, and Albrecht Schmidt. 2019. The Role of HCI in Reproducible Science: Understanding, Supporting and Motivating Core Practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3290607.3312905

[17] Jörg Fehr, Jan Heiland, Christian Himpe, and Jens Saak. 2016. Best practices for replicability, reproducibility and reusability of computer-based experiments exemplified by model reduction software. *AIMS Mathematics* 1, 3 (2016), 261–281. doi:10.3934/Math.2016.3.261

[18] Robert Gentleman and Duncan Temple Lang. 2007. Statistical analyses and reproducible research. *Journal of Computational and Graphical Statistics* 16, 1 (2007), 1–23.

[19] Sean Grant, Katherine S Corker, David Thomas Mellor, Suzanne LK Stewart, Aidan Cashin, Malgorzata Lagisz, Evan Mayo-Wilson, David Moher, Daniel Umpierre, Virginia Barbour, et al. 2024. TOP 2025: An Update to the Transparency and Openness Promotion Guidelines. *MetaArXiv* 1, 1 (2024), 1–12. doi:10.31222/osf.io/nmfs6_v2

[20] Tom Hardwicke, Maya Mathur, Kyle MacDonald, Gustav Nilsonne, Banks, Mallory Kidwell, Alicia Hofelich Mohr, Elizabeth Clayton, Erica Yoon, Michael Henry Tessler, Richie Lenne, Sara Altman, Bria Long, and Michael Frank. 2018. Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society open science* 5, 8 (2018), 180448.

[21] James J. Heckman. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47, 1 (1979), 153–161. http://www.jstor.org/stable/1912352

[22] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is once enough? on the extent and content of replications in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3523–3532. doi:10.1145/2556288.2557004

[23] Olga Iarygina, Kasper Hornbæk, and Aske Mottelson. 2025. Demand characteristics in human–computer experiments. *International Journal of Human-Computer Studies* 193 (2025), 103379. doi:10.1016/j.ijhcs.2024.103379

[24] John PA Ioannidis. 2012. Why science is not necessarily self-correcting. *Perspectives on Psychological Science* 7, 6 (2012), 645–654. doi:10.1177/1745691612464056

[25] John P. A. Ioannidis. 2014. How to Make More Published Research True. *PLOS Medicine* 11, 10 (10 2014), 1–6. doi:10.1371/journal.pmed.1001747

[26] Matthew Kay, Steve Haroz, Shion Guha, Pierre Dragicevic, and Chat Wacharamanotham. 2017. Moving Transparent Statistics Forward at CHI. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 534–541. doi:10.1145/3027063.3027084

[27] Matthew Kay, Cynthia Matuszek, and Sean A. Munson. 2015. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3819–3828. doi:10.1145/2702123.2702520

[28] Mallory C. Kidwell, Ljiljana B. Lazarević, Erica Baranski, Tom E. Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, Agnieszka Slowik, Carina Sonnleitner, Chelsey Hess-Holden, Timothy M. Errington, Susann Fiedler, and Brian A. Nosek. 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology* 14, 5 (05 2016), 1–15. doi:10.1371/journal.pbio.1002456

[29] Agnes Mercedes Kloft, Robin Welsch, Thomas Kosch, and Steeven Villa. 2024. "AI enhances our performance, I have no doubt this one will do the same": The Placebo effect is robust to negative descriptions of AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 299, 24 pages. doi:10.1145/3613904.3642633

[30] 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, 2 (2019), 408–429. arXiv:https://doi.org/10.1080/13658816.2018.1508687 doi:10.1080/13658816.2018.1508687

[31] Thomas Kosch, Robin Welsch, Lewis Chuang, and Albrecht Schmidt. 2023. The placebo effect of artificial intelligence in human–computer interaction. *ACM Transactions on Computer–Human Interaction* 29, 6, Article 56 (2023), 32 pages. doi:10.1145/3529225

[32] Randall J LeVeque. 2013. Top ten reasons to not share your code (and why you should anyway). *Siam News* 46, 3 (2013), 1–6.

[33] Robert K Merton. 1973. *The sociology of science: Theoretical and empirical investigations.* University of Chicago press, Chicago, USA.

[34] E. Miguel, C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn, and M. Van der Laan. 2014. Promoting Transparency in Social Science Research. *Science* 343, 6166 (2014), 30–31. arXiv:https://www.science.org/doi/pdf/10.1126/science.1245317 doi:10.1126/science.1245317

[35] Richard D Morey, Christopher D Chambers, Peter J Etchells, Christine R Harris, Rink Hoekstra, Daniël Lakens, Stephan Lewandowsky, Candice Coker Morey, Daniel P Newman, Felix D Schönbrodt, et al. 2016. The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. *Royal Society open science* 3, 1 (2016), 150547.

[36] Marcus R Munafò, Brian A Nosek, Dorothy VM Bishop, Katherine S Button, Christopher D Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J Ware, and John PA Ioannidis. 2017. A manifesto for reproducible science. *Nature human behaviour* 1, 1 (2017), 0021. doi:10.1038/s41562-016-0021

[37] 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. doi:10.1126/science.aab2374

[38] Brian A Nosek, Tom E Hardwicke, Hannah Moshontz, Aurélien Allard, Katherine S Corker, Anna Dreber, Fiona Fidler, Joe Hilgard, Melissa Kline Struhl, Michèle B Nuijten, et al. 2022. Replicability, robustness, and reproducibility in psychological science. *Annual review of psychology* 73, 1 (2022), 719–748.

[39] Michèle B. Nuijten, Marjan Bakker, Esther Maassen, and Jelte M. Wicherts. 2018. Verify original results through reanalysis before replicating. *Behavioral and Brain Sciences* 41 (2018), e143. doi:10.1017/S0140525X18000791

[40] 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, 2 (2020), 229–237. arXiv:https://doi.org/10.1177/2515245920918872 doi:10.1177/2515245920918872

[41] Roger D Peng. 2011. Reproducible research in computational science. *Science* 334, 6060 (2011), 1226–1227.

[42] Heather A. Piwowar, Roger S. Day, and Douglas B. Fridsma. 2007. Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLOS ONE* 2, 3 (03 2007), 1–5. doi:10.1371/journal.pone.0000308

[43] O James Reichman, Matthew B Jones, and Mark P Schildhauer. 2011. Challenges and opportunities of open data in ecology. *Science* 331, 6018 (2011), 703–705.

[44] J. Schneider, T. Rosman, A. Kelava, and S. Merk. 2020. (Re)Building Trust? Journals' Open Science Badges Influence Trust in Scientists. *PsyArXiv preprint* 1, 1 (August 2020), 1–32. doi:10.23668/PSYCHARCHIVES.3364

[45] Jürgen Schneider, Tom Rosman, Augustin Kelava, and Samuel Merk. 2022. Do Open-Science Badges Increase Trust in Scientists Among Undergraduates, Scientists, and the Public? *Psychological Science* 33, 9 (2022), 1588–1604. arXiv:https://doi.org/10.1177/09567976221097499 doi:10.1177/09567976221097499 PMID: 36001881.

[46] Jens Jakob WH Sørensen, Mads Kock Pedersen, Michael Munch, Pinja Haikka, Jesper Halkjær Jensen, Tilo Planke, Morten Ginnerup Andreasen, Miroslav Gajdacz, Klaus Mølmer, Andreas Lieberoth, et al. 2016. Exploring the quantum speed limit with computer games. *Nature* 532, 7598 (2016), 210–213. doi:10.1038/nature17620

[47] Daniel Stockemer, Sebastian Koehler, and Tobias Lentz. 2018. Data Access, Transparency, and Replication: New Insights from the Political Behavior Literature. *PS: Political Science & Politics* 51, 4 (2018), 799–803. doi:10.1017/S1049096518000926

[48] Victoria Stodden. 2015. Reproducing statistical results. *Annual Review of Statistics and Its Application* 2, 1 (2015), 1–19.

[49] Victoria Stodden, Marcia McNutt, David H Bailey, Ewa Deelman, Yolanda Gil, Brooks Hanson, Michael A Heroux, John PA Ioannidis, and Michela Taufer. 2016. Enhancing reproducibility for computational methods. *Science* 354, 6317 (2016), 1240–1241.

[50] 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* 115, 11 (2018), 2584–2589. arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1708290115 doi:10.1073/pnas.

1708290115
[51] Poorna Talkad Sukumar, Ignacio Avellino, Christian Remy, Michael Ann DeVito, Tawanna R. Dillahunt, Joanna McGrenere, and Max L. Wilson. 2020. Transparency in Qualitative Research: Increasing Fairness in the CHI Review Process. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3334480.3381066
[52] Patrick Vandewalle, Jelena Kovacevic, and Martin Vetterli. 2009. Reproducible research in signal processing. *IEEE Signal Processing Magazine* 26, 3 (2009), 37–47.
[53] Steeven Villa, Thomas Kosch, Felix Grelka, Albrecht Schmidt, and Robin Welsch. 2023. The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior. *Computers in Human Behavior* 146 (2023), 107787. doi:10.1016/j.chb.2023.107787
[54] Jan B. Vornhagen, April Tyack, and Elisa D. Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) *(CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 4–18. doi:10.1145/3410404.3414229
[55] Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. 2020. Transparency of CHI Research Artifacts: Results of a Self-Reported Survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376448
[56] Chat Wacharamanotham, Matthew Kay, Steve Haroz, Shion Guha, and Pierre Dragicevic. 2018. Special Interest Group on Transparent Statistics Guidelines. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3170427.3185374
[57] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.
[58] Julie Williamson, Jie Li, Vinoba Vinayagamoorthy, David A. Shamma, and Pablo Cesar. 2021. Proxemics and Social Interactions in an Instrumented Virtual Reality Workshop. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 253, 13 pages. doi:10.1145/3411764.3445729
[59] Max Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, and Jeffrey Nichols. 2012. RepliCHI SIG: from a panel to a new submission venue for replication. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI EA '12)*. Association for Computing Machinery, New York, NY, USA, 1185–1188. doi:10.1145/2212776.2212419
[60] Max L. Wilson, Ed H. Chi, Stuart Reeves, and David Coyle. 2014. RepliCHI: the workshop II. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 33–36. doi:10.1145/2559206.2559233
[61] Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. 2011. RepliCHI - CHI should be replicating and validating results more: discuss. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA '11)*. Association for Computing Machinery, New York, NY, USA, 463–466. doi:10.1145/1979742.1979491
[62] Max L. L. Wilson, Paul Resnick, David Coyle, and Ed H. Chi. 2013. RepliCHI: the workshop. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 3159–3162. doi:10.1145/2468356.2479636
[63] Burak Yildiz, Hayley Hung, Jesse H. Krijthe, Cynthia C. S. Liem, Marco Loog, Gosia Migut, Frans A. Oliehoek, Annibale Panichella, Przemysław Pawełczak, Stjepan Picek, Mathijs de Weerdt, and Jan van Gemert. 2021. ReproducedPapers.org: Openly Teaching and Structuring Machine Learning Reproducibility. In *Reproducible Research in Pattern Recognition: Third International Workshop, RRPR 2021, Virtual Event, January 11, 2021, Revised Selected Papers*. Springer-Verlag, Berlin, Heidelberg, 3–11. doi:10.1007/978-3-030-76423-4_1

A Guidelines on how to make a reproducible repository

These guidelines can be used either as a step-by-step instruction when creating a new repository or as a checklist for assessing and improving the reproducibility of existing ones. For researchers who are new to open science, we provide an introduction to repositories and practical advice on how to set them up.

Setting up the repository

(1) **Create an open science repository.** The first step is to make your data and code accessible in a proper repository. Repository is a place where you can store your code, data, and any related research materials. Repositories compliant with the FAIR principles include OSF and Zenodo. Some people working with version control prefer to use GitHub. If this is the case, a GitHub repository can be archived on Zenodo to make it persistent and immutable, or it can be connected to OSF (https://help.osf.io/article/211-connect-github-to-a-project). We would recommend using OSF for its simplicity and integration with other tools. However, Zenodo is better suited for storing large files.
Helpful getting-started guides for those new to repositories:
GitHub: https://docs.github.com/en/repositories/creating-and-managing-repositories/quickstart-for-repositories
OSF: https://help.osf.io/article/353-welcome-to-projects

(2) **Create a README or Wiki page.** A README or Wiki is usually the first item a visitor sees, and it should make the repository understandable and easy to navigate. A good README often contains:
   (a) A description of what the project is about.
   (b) The paper to which it belongs.
   (c) The structure of the repository.
   (d) A description of what each file in the repository does.
   (e) The requirements for running the code.
   (f) The location of the working directory.
   (g) Instructions on how to get started (for example, the order of script execution).
   (h) Information on where users can ask for help.

   Additional resources for structuring a README file:
   GitHub README guide: https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-readmes
   OSF Wiki guide: https://help.osf.io/article/389-wiki
   Good Wiki example: https://osf.io/8q7t6/wiki/home/ [29]

Data and documentation

(3) **Include requirements for analysis.** Reproduction often fail because the required software environment is unclear. Always list the packages and libraries used. A good practice is to specify package versions. To get the versions of the packages used in R, use `>sessionInfo()` command. In Python, use `>conda list` if you work with conda, or `>pip freeze` if you work with pip. Alternatively, you can use `sessioninfo/version-information` packages.

(4) **Organize your files.** It is handy when data and analysis scripts are located in different folders, for example `data` and `analysis`. If you have multiple studies in your repository, it is a good practice to make a separate folder for each. OSF allows users to create subprojects within a project, and some authors use subprojects for separating studies. Counterintuitively, we would not recommend doing this, as it makes finding the right files more complicated.

(5) **Name files meaningfully.** A good file name self-explains the content of that file. Examples of the names could be: `survey_data.csv`, `exploratory_analysis.Rmd`.
(6) **Include and document your data.** The most common reason for reproduction failure is missing data files. Ensure all necessary data is included. Give columns meaningful names such as `age`, `id`, `[questionnaire]_q1`. Check that file names in the code and in the repository match, and that the paths are corrects. If your data has a specific format, provide instructions for how to open or use it. Make sure data is anonymized. If some data cannot be shared, specify this clearly in the README.
(7) **Create a codebook.** A codebook is essential for understanding what the data means. At minimum, it should include each column name and its meaning. Ensure that variable names in the codebook match those in the data. A more detailed data dictionary can also specify allowed values and measurement units. A detailed instruction for creating data dictionary can be found here: https://help.osf.io/article/217-how-to-make-a-data-dictionary

Code practices

(8) **Clean and annotate your code.** Well-annotated code helps others (and your future self) understand what is going on. Some tips we recommend:
- Call all libraries at the beginning of the script.
- If using object-oriented programming, describe which scripts need to be run and how they interact.
- If you define your own custom function, explain what it is supposed to do. This is especially helpful when the code needs troubleshooting.
- Divide the code into logical sections, such as "Data cleaning" or "Statistical analysis."
- If using Jupyter Notebook or RMarkdown, clear the workspace at the top of each script.
- If the script generates files or figures, specify what are the expected outputs.

Tools like Copilot or AI can assist with annotations, but always double-check their accuracy.
(9) **Help others reuse your code.** Specify how scripts should be executed, which commands are required, which order of execution should be followed, and which software is needed. For RStudio users, sharing an `.Rproj` file is useful, as it manages the relationships between files automatically.
(10) **Help readers understand your code.** If using Jupyter Notebooks or RMarkdown, provide a knitted HTML or PDF of the executed notebook. This allows others to compare your original output with their reproduction attempt, and makes it easier to track down potential errors. If the code produces figures or files outside a notebook, create a folder containing expected outputs for reference.
(11) **Connect your code output to the paper.** It helps a lot if users can directly see how your code outputs relate to the paper. This can be done by:
- Explaining the link between scripts and parts of the paper in the README.
- Adding headers or annotations describing sections in the paper in the code (as in here [58]).
- Naming scripts according to sections of the paper.
- Naming outputs after parts of the paper (e.g., `figure1.png`).
- Writing the full paper as a reproducible document with embedded code chunks, for example using Quarto or RMarkdown. This way, the analysis, figures, and results are generated together with the text (as in here [27]).

(12) **Include all the relevant code to the repository** Add all the code you used to generate the results of the paper. Helpful guides for adding files can be found here:
OSF: https://help.osf.io/article/387-project-files
GitHub: https://docs.github.com/en/repositories/working-with-files/managing-files/adding-a-file-to-a-repository

Verification and publishing

(12) **Test reproducibility.** After setting up your repository, download it and rerun all code from scratch. This ensures that everything works as expected. Ideally, ask a colleague to test it as well. Even better, run it on a different operating system, since OS differences can cause errors.
(13) **Make your repository public.** After the review cycle, do not forget to make your repository public:
OSF: https://help.osf.io/article/285-control-your-privacy-settings#Make-your-project-or-components-public
GitHub: https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/managing-repository-settings/setting-repository-visibility
(14) **Attach the repository link to your paper.** Always include a link to the repository in your paper to ensure it is findable. To increase discoverability, add metadata on OSF: https://help.osf.io/article/572-metadata-to-increase-discoverability
To make the repository easier to cite, create a persistent digital object identifier (DOI).
OSF instructions: https://help.osf.io/article/575-persistent-identifiers-pids
If you use Github, archive the repository and issue a DOI via Zenodo: https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content.
(15) **Document changes.** If you update the repository after publication, clearly specify these changes in the README or Wiki file. Without documenting post-publication changes, the repository may diverge from what was described in the paper, making it unclear which version supports the published results.

Additional recommendations

- Plan to share your materials from the beginning. Planning in advance helps to organize the files and code on the go, reducing the time needed for preparing materials.
- Comment the code as you go. It will help yourself to remember what you have done.
- To ensure fully reproducible environment, consider using tools like Docker, or ReproZip. They enable to capture the exact software environment so others can reproduce your

results even years later. However, it might take some time to learn how to use them.
- Whenever possible, share data in non-proprietary formats (e.g., `.csv`, `.txt` . This avoids lock-in and makes data accessible for more people.
- A detailed README and clear comments helps others to reuse your work. Reproducibility is as much about communication as it is about technology.
- A good practice to learn reproducibility is to try to reproduce someone's work. It will quickly highlight what is helpful and what is frustrating.