"Good enough" model reproducibility: A practical guide for early career scholars in socio-environmental modelling

Takuya Iwanaga[1,2,3*], Jazmin Zatarain Salazar[4,3], Nianhua Liu (刘年华)[5,3], Zhiyi Zhu (朱之一)[6,3], Radmila Jovanović[7,8,3], Liang-Jun Zhu (朱良君)[9,3], Hugo Salinas[10,3], Ismael Kherroubi Garcia[11,3], and Tatiane Micheletti[12,13,14, 3]

[1]Institute for Water Futures, Australian National University, Canberra, Australia
[2]Australian Institute of Marine Science, Townsville, Queensland, Australia
[3]Open Modeling Foundation, OMF
[4]Hippo DAI Lab, Delft University of Technology, Delft, Netherlands
[5]Chair of Cartography and Visual Analytics, Technical University of Munich, Munich, Germany
[6]Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing, Jiangsu, China
[7]Institute of Agricultural Economics, Belgrade, Serbia
[8]Department of Geography, Faculty of Tourism, University of Malaga, Malaga, Spain
[9]Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[10]School of Biological Sciences, University of Western Australia, Perth, Western Australia, Australia
[11]Kairoi, AI Ethics & Research Governance
[12]TU Dresden, Dresden, Germany
[13]Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany
[14]German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany

Abstract

The expectation of "perfect" research reproducibility often acts as a deterrent rather than a catalyst for Early Career Scholars (ECSs) navigating the challenges of academia. Caught between the technical overhead of idealized workflows and the institutional pressure for publication velocity, ECSs require a pragmatic strategy that balances scientific rigor with the reality of finite resources. This paper provides that middle ground through a "good enough" framework for socio-environmental modelling, grounded in a fitness-for-purpose logic. We conceptualize reproducibility as a spectrum of five stages of computational reproducibility maturity, ranging from exploratory private sandboxes to reusable tools. Within this continuum, we identify Stage 3 (Portable Analysis) as the achievable and sufficient baseline for standard peer-reviewed research. To bridge the gap between theory and execution, we detail five foundational pillars – Documentation, Organization, Version Control, Self-Replication, and Sharing – and offer a tiered pre-submission checklist to help researchers verify their work. While providing ECSs with a technical roadmap, we emphasize that resolving the reproducibility crisis requires systemic shifts in how institutions and funding agencies resource the labor of transparency. Ultimately, this guide establishes a sustainable professional baseline for a new generation of modelers to produce work that is transparent, credible, and sustainable.

Keywords

Reproducibility; early career scholars; computational research; best practices; workflows

Correspondence:
* Contact T. Iwanaga at t.iwanaga@aims.gov.au

Cite this article as:
Iwanaga, T., Zatarain Salazar, J., Liu, N., Zhu, Z., Jovanović, R., Zhu, L.-J., Salinas, H., Garcia, I.K. & Micheletti, T.
"Good enough" model reproducibility: A practical guide for early career scholars in socio-environmental modelling
Socio-Environmental Systems Modelling, vol. 8, 19054, 2026, doi:10.18174/sesmo.19054

This work is **licensed** under a Creative Commons Attribution-NonCommercial 4.0 International License.



Socio-Environmental Systems Modelling
An Open-Access Scholarly Journal
http://www.sesmo.org

1. Introduction

Reproducibility is the cornerstone of scientific research, yet for Early Career Scholars (ECSs), the transition from principle to practice is often fraught with structural hurdles. While the "reproducibility crisis" is a global concern (Baker, 2016), the burden of solving it falls disproportionately on those with the least institutional power and the most precarious job security. ECSs, such as doctoral students and post-docs, must navigate a system that rewards publication velocity over the long-term rigor of computational workflows (Melsen et al., 2017; Micheletti et al., 2024). Because these pressures are systemic, reproducibility advice cannot be a 'one-size-fits-all' directive. We treat ECSs as a distinct group with specific needs, rather than simply as junior researchers on a shared continuum, because their position within the research system is defined by several structural conditions that make traditional reproducibility advice especially difficult to implement.

First, high institutional mobility creates frequent discontinuities in research materials. As ECSs move between short-term positions – often across different institutions and research groups – code and data are frequently left behind or insufficiently documented for successors, a process that effectively 'orphans' scientific output. The development of this manuscript is itself illustrative of these pressures: revisions were delayed by approximately two years as a direct consequence of authors relocating institutions and balancing the voluntary nature of this work against competing project commitments. Second, ECSs are disproportionately shaped by the social and institutional context of their current lab (Melsen et al., 2019; Wirtz and Nowak, 2017). Moving between groups can expose researchers to conflicting standards, yet the power asymmetry inherent in academia often leaves ECSs with the least capacity to influence their working conditions or advocate for the time required to implement rigorous workflows. Finally, this stage is the formative period during which research habits and methodological norms are established (Kohrs et al., 2023). Because practices adopted early in a researcher's career tend to persist throughout it, providing a specific 'how-to' guide for ECSs offers the greatest leverage for long-term cultural change in science.

Recognizing these constraints, this paper provides a practical guide to 'good enough' reproducibility. We argue that reproducibility is not a binary 'pass/fail' requirement but a fit-for-purpose strategy (Hamilton et al., 2022), where the "quality assurance effort should be proportionate in response to the risks associated with the intended use of the analysis" (Edmonds, 2016; HM Treasury, 2015), i.e., where the level of technical rigor is matched to the project's specific goals. By navigating four distinct stages of computational maturity, from the Private Sandbox to the Robust Model, ECSs can make informed choices about where to invest their limited time. While our advice is contextualized for socio-environmental modelling, these principles are designed to help all computational researchers produce work that is transparent, credible, and, crucially, reusable by their future selves.

1.1 What is reproducibility?

At its core, science is built on the proposition and testing of hypotheses. As Popper (1959) argued, scientific observations must occur with repeated regularities; ideally, anyone must be able to repeat a test to ensure that findings are not mere coincidences. In the computational age, this principle is operationalized through reproducibility: the ability of researchers not involved in an original study to arrive at identical results by recreating the same analytical process (Committee on Reproducibility and Replicability in Science et al., 2019).
While the terms reproducibility and replicability are often conflated, they represent different scales of verification. To some, reproducibility is the foundation on which replicability is enabled (Buckheit and Donoho, 1995); to others, they are separate concerns (Plesser, 2018). We categorize these concepts into foundational practices – transparency and accessibility – which enable scientific rigor, and verification experiments – reproducibility and replicability – which explicitly test it. Table 1 provides our framework for these distinctions based on the anchor of the research and the components being tested.

Table 1: Dimensions of scientific rigor and key terms related to research reproducibility and their distinctions. These terms are frequently conflated in literature but represent distinct dimensions of the research lifecycle. This guide distinguishes between foundational practices (transparency and accessibility), which are necessary conditions for rigor, and verification experiments (reproducibility and replicability), which test the stability of results based on different anchors of data, methods, and the computational environment.

Concept	Nature	Goal	What stays constant?	What is tested?
Transparency	Policy and practice	Ensure process clarity	The research process	The clarity of documentation of methods, decisions and rationale
Accessibility	Policy and practice	Ensure retrieval or use	The research materials	Availability, format, & license
Reproducibility	Experiment	Computational integrity	Original data, code, & methods	Computational environment
Replicability	Experiment	Scientific reliability	Research question or hypothesis	New data, new investigators

In socio-environmental modelling, we must also consider numerical reproducibility (or replicability, depending on the definition), which requires the explicit fixing of "random seeds" to ensure that stochastic processes yield identical results across runs. However, our focus here is not on achieving identical results across all possible hardware configurations; rather, we emphasize a pragmatic standard where the research process is sufficiently transparent and portable for others to audit the logic and recreate the findings with reasonable effort.

We treat transparency and accessibility as necessary conditions for reproducibility and use "reproducibility" as an umbrella concept. At a “micro” level, it speaks to the trustworthiness of an individual study; at a “macro” level, it provides the foundation on which further scientific progress is made (Hutton et al., 2016). Ultimately, a lack of reproducibility can diminish the achievements of entire fields of study (Open Science Collaboration, 2015) and is vital for the development of effective government policy (Saltelli & Funtowicz, 2017).

Beyond these foundational benefits, we identify three distinct drivers for ECS engagement with research reproducibility. First, we currently exist in an era of increasing skepticism toward scientific authority. Because reproducibility increases trust, the burden falls on ECSs to adopt and spread methods that ensure research is verifiable. By moving toward a paradigm where open research is standard practice, the influence of irreproducible research on government policy can be diminished (Saltelli, 2018; Saltelli & Funtowicz, 2017), while the perception of 'bad science' or academic misconduct is avoided through the transparency of data and methods (Higginson & Munafò, 2016).

Second, a culture of openness, evidenced by sharing code, data, and knowledge, supports “scientific hitchhiking” (Micheletti et al., 2024), increasing the velocity of scientific discovery. By providing transparent and accessible workflows, ECSs promote open collaboration and set a benchmark that prevents the stagnation of entire fields (Open Science Collaboration, 2015). In this context, reproducibility is not merely a hurdle but a mechanism for making scientific progress more efficient (Theiler, 1990).

Finally, as research becomes increasingly complex and interdisciplinary, reproducibility allows for a compositional workflow. Data and code are no longer isolated outputs but serve as 'blocks' to build toward new studies (Berger et al., 2024; Filatova et al., 2025; Kondyli et al., 2024; Samuel & König-Ries, 2021). This underscores the necessity of the FAIR principles (Findable, Accessible, Interoperable, and Reusable), ensuring that research remains valuable and reusable in later investigations long after the initial publication (Barton et al., 2022b, 2022a).

1.2 The reproducibility crisis

While the FAIR principles provide a roadmap for research longevity, their widespread adoption is constrained by a systemic 'reproducibility crisis' that stems from a blend of external incentives and internal research cultures (Micheletti et al., 2024). The structural barriers facing individual ECSs are symptoms of an academic landscape that often prioritizes rapid publication over methodological rigor. Frequent institutional transitions fragment research continuity; power asymmetries discourage or marginalize advocacy for better practices; and short

funding cycles (Meirmans et al., 2019) force publication timelines that leave little room for code quality or accessibility. Many publishers still lack mandatory reproducibility standards, and the review system often overlooks the value of repeated research (Schnell, 2018). Furthermore, professional recognition rarely rewards open science; an analysis of "Best Researcher" awards in ecology and evolution found that only one mentioned open practices (Lagisz et al., 2023), although some punctual initiatives which consider open practices and reproducibility (i.e., Community Surface Dynamics Modeling System, 2026) may suggest a slow increase in recognition of open practices. The lack of incentives contributes to a cycle where critical discrepancies go unnoticed, leading to a bias toward positive results (Higginson & Munafò, 2016; Leek & Jager, 2016; Saltelli et al., 2019) or even academic misconduct (Enserink, 2021; Esbaugh, 2023).

The consequences of irreproducibility are not abstract. In environmental and earth system sciences, the sensitivity of model outputs to software version, numerical implementation, and parameter choices is well-documented (Bhandari Neupane et al., 2019). Melsen et al., (2019) showed that seemingly minor differences in the implementation of the same hydrological model across research groups produced variations large enough to alter conclusions. More broadly, the controversy over fish behavior responses to ocean acidification – where a body of published results could not be independently reproduced – illustrates how irreproducible findings can distort research agendas for years (Enserink, 2021; Esbaugh, 2023). These incidents reflect systemic vulnerabilities in research workflows where process-based models risk becoming "black boxes" (Seppelt & Richter, 2005).

This challenge is especially acute in socio-environmental modelling, which requires navigating climate change, deep uncertainty, and interdisciplinary integration (Grimm, 2023; Jebeile & Crucifix, 2020; Schuwirth et al., 2019). While standards for documentation and interoperability have been developed (Benz et al., 2001; Hutton et al., 2020), their uptake remains haphazard. A cultural preference for creating new models from scratch, rather than leveraging established components, persists (Hutton et al., 2016; Jenkins et al., 2023; Melsen et al., 2017; Pellizzari et al., 2017), hindering the transparency essential for scientific progress.

Ultimately, ECSs are the group most affected by these failures, yet they are often the least equipped to address them unilaterally. Expecting full reproducibility from ECSs without institutional support places an unrealistic burden on those with the fewest resources. We propose a resolution through a fitness-for-purpose framing (Hamilton et al., 2022), in a tiered approach. Recognizing that the appropriate standard of reproducibility is not a universal constant, but depends on the purpose, scope, and audience of the work, enabling ECSs to make deliberate choices about where to direct their effort.

2. The five stages of reproducibility

To operationalize fitness-for-purpose in a tiered approach, we conceptualize reproducibility maturity as a progression through five distinct stages (Figure 1). These stages begin with (1) a private sandbox, which can then be (2) transparently archived, (3) made into a discrete set of data and code for portable analysis, (4) containerized into a Robust Model Archive or (5) developed into a reusable tool for others to adopt. These five stages represent increasing levels of investment in formalization and accessibility. Rather than a strict hierarchy to be ascended for every project, these stages offer a menu of options where the “good enough” standard is determined by the specific context of the work.

To help ECSs assess which stage is appropriate for a given project, we identify four contextual dimensions:

1. **Audience and Reuse**: Is the code intended for personal use, a small group of collaborators, or the wider public?
2. **Longevity**: Does the analysis need to be functional for a few months, or must it remain executable for several years?
3. **Policy Relevance and Risk**: What are the consequences of an error? Does the model inform high-stakes management decisions or generate preliminary hypotheses?
4. **Complexity and Collaboration**: Is it a single-script analysis or a multi-institutional, interdisciplinary modeling effort?

By evaluating a project against these dimensions, ECSs can make an informed choice between the following stages.

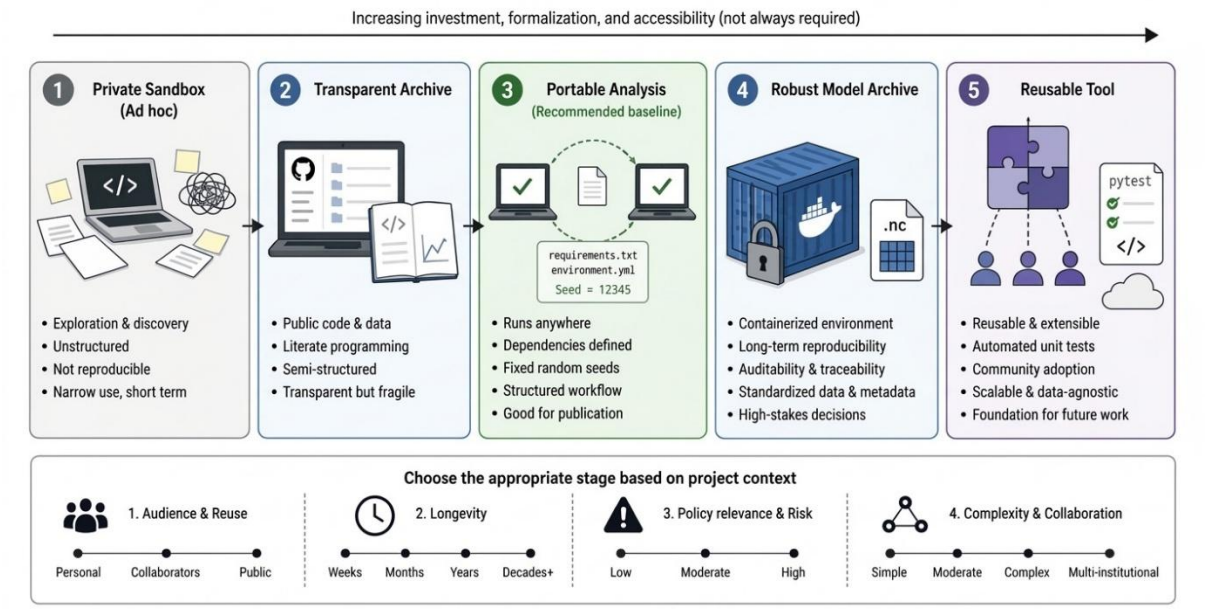


Figure 1: Conceptual framework of reproducibility maturity as a fitness-for-purpose continuum. Research workflows can range from an ad hoc private sandbox (Stage 1) to a fully reusable, community-oriented tool (Stage 5), with increasing levels of formalization, transparency, and technical investment. Rather than representing a strict hierarchy, these stages provide a menu of options. The appropriate level of reproducibility is determined by four contextual dimensions: audience and reuse, longevity, policy relevance and risk, and complexity and collaboration. For most peer-reviewed studies, a portable and well-documented workflow (Stage 3) provides a practical and sufficient baseline.

Stage 1: The private sandbox (ad hoc)

Every research project begins with a period of discovery. This stage is captured by an anecdote from Nobel laureate Albert Szent-Györgyi who, when asked how he could justify basic research without knowing the outcome, replied: “I don’t know, that is why it is research” (Huxley et al., 1994). This stage is characterized by a minimalist structure that often lacks consistency, with code, data, and information scattered in a disorganized manner. Research is inherently "messy", and ECSs often undergo rapid cyclic iterations of exploration and learning to navigate unknowns (Jakeman et al., 2006). This iterative process akin to "agile" workflows (Iwanaga et al., 2021b) frequently produces a significant amount of “mess”, including unfinished scripts, intermediate data in proprietary formats, and confusing file names such as *notebook_FINAL_v2_ACTUAL_old.ipynb*.

While the "Sandbox" stage is essential for intellectual freedom and “failing fast” to allow rapid learning, exploration and hypothesis testing, it is not reproducible in and of itself. For small exploratory work, it may be sufficient to communicate a trial-and-error process to a close collaborator, but it is never sufficient for formal publication. Relying on an ad hoc approach for published results renders the work inaccessible even to the original author after only a few months (Barton et al., 2022a). This stage is defensible only when the audience is narrow, policy relevance is low, and the longevity of the analysis is intentionally short.

Stage 2: The transparent archive

A more semi-structured approach is adopted in this stage, with code organized into folders that separate data processing from analysis and adopting naming conventions, such as formatting dates as YYYY-MM-DD, to avoid confusion between international collaborators who use conflicting date formats. Some researchers may also start to utilize "literate programming" (Knuth, 1984) through computational notebooks like Jupyter Notebooks (https://jupyter.org/), Quarto (https://quarto.org/), RMarkdown (https://rmarkdown.rstudio.com/), and MATLAB’s Live Scripts (https://mathworks.com/help/matlab/matlab_prog/create-live-scripts.html). These environments allow text, imagery, and code to be mixed into a computational document that makes the analytical flow accessible to researchers from various backgrounds (Ayllón et al., 2021; Kim et al., 2018). When thinking about publications, researchers also start moving toward transparency by making their materials public (e.g., via GitHub or Zenodo), often to satisfy journal mandates or institutional transparency norms. However, a larger adoption of these practices is still research-field-related.

However, while this stage provides a digital specimen of research, the workflow remains brittle. Because it lacks specific information on software dependencies and environment configurations, the code will likely not run on another machine without significant manual troubleshooting. This stage is good enough for low-complexity studies where the primary goal is for peers to audit mathematical logic or equations, rather than to execute the model themselves. It represents the minimum recommended standard for any work accompanying a journal submission that describes the application of models and methodologies.

Stage 3: The portable analysis

As workflows become more consistent, they grow in terms of portability, which ensures the work can survive the high institutional mobility characteristic of an ECS. This is the stage where the distinction between "should" and "must" becomes a functional reality. Portable projects adopt established conventions for organization, such as the Cookie-cutter data science structure (https://cookiecutter-data-science.drivendata.org/), and ensure that all hard-coded file paths are removed so the code can execute on any computer (Donoho, 2024).

To mitigate the risks of software environment-induced irreproducibility, researchers at this stage utilize dependency management tools to explicitly document the software stack and library versions, typically through configuration files such as *requirements.txt* or *environment.yml* for Python. Crucially, for stochastic models, this stage mandates the fixing of random seeds. This ensures that the specific sequence of random numbers is identical across runs, making the numerical results as reproducible as possible. There are, however, limits to numerical reproducibility. Even with a fixed seed, results may differ across operating systems, hardware architectures, compilers, or software library and language versions. Differences may occur due to changes in the underlying random number generator, compiler optimizations, or simply the order of parallel operations (Melsen et al., 2019; Seppelt & Richter, 2005). For these reasons, fixing random seed values is strongly recommended to support reproducibility within a given research context. Precise numerical equivalence across environments is a secondary concern relative to reproducibility of model behavior and outcomes.

To ensure conceptual transparency, Stage 3 involves the use of standardized model description protocols, such as the ODD (Overview, Design concepts, and Details) protocol (Grimm et al., 2020, 2006). This provides a consistent structure for communicating model logic and assumptions to the scientific community, making the research "good enough" for peer-reviewed publication. For most ECSs, Stage 3 is the recommended standard for peer-reviewed manuscripts; it provides a practical threshold that balances scientific credibility with the finite time and technical resources available during a doctoral or postdoctoral program.

Stage 4: The robust model

The fourth stage, the robust model archive shifts the focus from simple portability toward the long-term preservation and auditability of a specific scientific study. This level of investment is warranted when the modeling effort informs high-stakes policy or management decisions, where the exact results must remain verifiable and transparently auditable decades after the initial publication. To achieve this degree of permanence, researchers utilize Containerization technologies, such as Docker or Singularity, to freeze the entire computational environment. This process encapsulates not just the scripts and data, but the specific version of the operating system and the entire library stack, effectively mitigating the risk of software decay and dependency challenges.

Data management at this stage also moves toward rigorous standardization, utilizing formats like NetCDF (James et al., 2022) to package multidimensional data alongside its associated metadata in a single, human-readable file. Furthermore, the conceptual documentation matures from describing the model toward comprehensive traceability frameworks like TRACE (Ayllón et al., 2021). This provides a meticulous narrative of the entire modeling cycle, including calibration procedures, evaluation metrics, and sensitivity analyses. In this state, the research exists as a bit-for-bit reproducible digital specimen: a static historical record that is immune to the changing landscape of modern computing, ensuring that the evidence base for critical decisions remains open to scrutiny.

It is important to acknowledge that the comprehensive application of these practices – particularly the implementation of advanced workflows, documentation standards, and model-sharing frameworks –often requires an investment of time and resources that exceeds the typical duration of PhD or postdoctoral projects; nevertheless, familiarizing oneself with these tools is a crucial first step for ECSs to begin aligning their research with the evolving standards of the field.

Stage 5: The reusable tool
The final stage, the reusable tool, represents the highest level of computational maturity, where the primary objective shifts from reproducing a specific past result to enabling the reuse, extension, and collaborative evolution of the model by the wider community. This leap requires the model to transition from a static, archived artifact into a flexible, data-agnostic software tool designed for community adoption. Because the code is intended to be modified or applied to novel datasets by external researchers, this stage mandates the implementation of Automated Unit Tests (utilizing frameworks like PyTest or Test.jl). These tests act as a fundamental safety net, automatically verifying that the core mathematical and physical logic remains intact even as the code is updated, refactored, or integrated into larger modeling frameworks. Tests should aim for a high percentage of code coverage.

Furthermore, this stage adopts modern, cloud-optimized data standards such as Zarr or Cloud Optimized Geotiffs (Barciauskas et al., 2023) to support large-scale, distributed computing and high-performance interdisciplinary collaboration. By separating the model logic from specific data inputs and providing automated verification, researchers ensure that their work can serve as robust and reusable building blocks (Berger et al., 2024) for future investigations and collaborative model extension. While this represents the most significant investment of time and expertise, it is the professional standard required for community modeling frameworks and software packages intended to serve as a lasting foundation for the socio-environmental modeling field.

2.1 Operationalizing reproducibility

To operationalize this framework, researchers are encouraged to evaluate their project's needs across five contextual dimensions and match them to the appropriate stage of computational maturity. The relationship between these stages and the technical practices later described in Section 3 is one of progressive intensity. A single practice – such as version control – may be adopted early for convenience, but its implementation becomes more rigorous as a project moves toward higher stages; for instance, while simple code hosting may suffice for Stage 2, Stage 4 requires the use of Git "tags" or "releases" to ensure a permanent, auditable link between the code and its results.

Table 2 serves as a roadmap for this process. It functions as both a decision matrix for project planning and a technical checklist for researchers to determine when their work has reached a "good enough" state for its intended purpose. It also connects the five stages of "good enough" to the practice focus described in the next section.

The five stages defined above characterize reproducibility as a spectrum of technical formalization. The specific practices described in the following section – Documentation, Organization, Version Control, Self-Replication, and Sharing – serve as the concrete mechanisms through which a researcher moves between these stages. Specifically, the transition from an exploratory Sandbox (Stage 1) to a Portable Analysis (Stage 3) is primarily achieved through systematic documentation and logical organization (Practices 3.1–3.2). Moving beyond this to reach a Robust Archive (Stage 4) or a Resilient Tool (Stage 5) requires the adoption of version control, self-replication protocols, and standardized community frameworks (Practices 3.3–3.5).

3. Practices for reproducibility

The five stages described in Section 2 characterize reproducibility as a destination; the practices described here are the mechanisms for reaching it. No single practice is sufficient in isolation, and their value compounds: documentation without organization produces a well-described but unnavigable project, while version control without documentation leaves a complete but uninterpretable history. For the ECS, the challenge is not just how to implement these practices, but how to do so efficiently within the time constraints of a research program. By consistently integrating these five pillars into their daily research activities, ECSs can significantly improve the reliability of their work while building the professional habits necessary for long-term career success. For those interested in a short, amusing perspective on the pitfalls of doing the opposite, we recommend Hong et al. (2015) as a cautionary guide on how to make research irreproducible.

Table 1: Decision matrix for selecting the appropriate reproducibility stage based on five contextual dimensions. Researchers are encouraged to use this table as a starting point for deliberate reflection rather than a mechanical scoring exercise.

Dimension	Stage 1: Sandbox	Stage 2: Archive	Stage 3: Portable	Stage 4: Robust	Stage 5: Resilient
Primary Goal	Rapid prototyping; failing fast	Transparency of mathematical logic	Peer execution and portability	Long-term auditability and permanence	Collaborative reuse and extension
Target Audience	The individual researcher	Peers auditing equations/logic	Peer-reviewers and active collaborators	Policy-makers and future auditors	The wider modeling community
Expected Longevity	Weeks (short-term)	Project duration (1–2 years)	1–3 years (post-publication)	Decades (long-term archive)	Persistent (evolving tool)
Policy/Risk Relevance	None; exploratory work	Low; descriptive research	Moderate; peer-reviewed findings	High; informs management/policy	Foundational; community block
Complexity/ Collaboration	Solo effort	Single-author or small team	Standard interdisciplinary team	Cross-institutional collaboration	Multi-community, global effort
Documentation Standard	Personal notes/lab diary	Researcher-defined README; data provenance	Standardized description (e.g., ODD protocol)	Formal protocol (e.g., TRACE)	Full protocol + API documentation
Code Organization	Messy scripts; hard-coded paths	Logical folder structure; commented code	Relative paths; style guide; fixed random seeds	Domain-specific conventions; linting applied	Modular design; automated style checks
Data Standards	Proprietary/ Custom formats	Human-readable; documented metadata	Consistent naming; metadata standards	Open standards (e.g., NetCDF)	Cloud-optimized (e.g., Zarr, COGS)
Version Control	None	Repository used for hosting/sharing	Systematic commits; documented changes	Tagged releases; permanent audit trail	CI/CD integration; DOI assignment
Licensing	None	Generic/ Undefined	Open license (e.g., MIT, GPL) specified	Consistent with community standards	Permissive; multi-author ready
Verification	Memory/ Intuition	None	Author self-replication attempted	Independent verification/ Collaborator audit	Automated Unit Test suite
Key Technical Requirements	None formal	Literate Programming (Jupyter/RMD)	Dependency files (requirements.txt) ; Readme	Containerization (Docker/ Singularity)	Automated Tests; Data-agnostic logic
"Good Enough" for...	Internal notes only; never for publication	Supplemental scripts for simple analyses	Standard journal publications	Policy-relevant models or high-impact studies.	Modeling frameworks and community software
Practice Focus (Section 3)	Internal iteration	3.1 Documentation	3.1 & 3.2 Organization	3.3 & 3.4 Versioning & Verification	3.5 Sharing & Testing

3.1 Document, document, document

Clear, comprehensive documentation is the bridge that allows others to understand, replicate, and extend a piece of work. Within our framework, the lack of documentation is the primary signifier of a Stage 1 (Sandbox) project. Documentation must cover all facets of research, including project objectives, methodologies, data sources, and software dependencies. For a “good enough” standard for publication (Stage 3), researchers are encouraged to maintain, at a minimum, a README file. At a minimum, this file should detail the directory structure, the provenance of all datasets, and the specific software versions required for execution. This investment in clarity is supported by Wang et al. (2023), who found a significant correlation between the comprehensiveness of a README and the engagement a repository receives from the scientific community.

While popularity is not a direct proxy for scientific correctness, it reflects the accessibility of the work; therefore, prioritizing documentation is a strategic necessity for ECSs aiming to maximize the reach and long-term impact of their contributions.

Beyond technical details, documentation should record the rationale behind decisions. In software design, this is known as Architectural Decision Records (Kruchten, 2008; Rekha V & Muccini, 2018); in research, it involves explaining why specific methodologies were chosen or why the research plan pivoted. Recording these reflections, including limitations and potential biases (Ayllón et al., 2021; Kohrs et al., 2023), provides valuable context for future users. In lab-based research, maintaining a continuous record (much like a daily lab notebook) offers immediate benefits to the ECS in tracking the origin, nature, and processing of data. Finally, for models intended for publication (Stage 3 and above), documentation should adopt standardized protocols such as ODD to ensure the model's conceptual logic is as transparent as its code.

3.2 Organize code and data and make it coherent

Coherent organization ensures that all project components are accessible and understandable. A well-structured project simplifies navigation by separating raw data, processed data, code, and documentation. Adopting descriptive naming conventions and using relative paths is essential for computational portability, ensuring the workflow remains functional when moved between institutions. Furthermore, writing clean, well-commented code is necessary to explain algorithmic logic and specific execution requirements (McConnell, 2004; Wilson et al., 2014).

Apart from model-metadata, which documents the model purpose, assumptions, inputs, outputs, parameters, dependencies, execution requirements and applicable contexts, computing-environment metadata is a critical consideration (Zhu et al., 2023). As research data availability declines rapidly with article age (Vines et al., 2014), data and accompanying metadata must be provided in at least human-readable formats (Leipzig et al., 2021). Researchers can leverage established conventions from the Data Carpentry project (Wilson et al., 2021) and its Reproducible Research workshops (Data Carpentry Committee, 2016). While advanced approaches to record the entire computing environment exist (Brinckman et al., 2018), and domain-specific best practices continue to evolve (Jenkins et al., 2023; Schnell, 2018), the core objective remains a transparent and navigable project structure.

3.3 Use version control

Modern research is increasingly collaborative, interdisciplinary, and often conducted by decentralized teams (Iwanaga et al., 2021b, 2021a). Version control is currently the only viable code management tool that supports coordinating code changes (Sletholt et al., 2012). It serves as a detailed log of the work's progression, where each commit explains what changed and why. This transparency significantly enhances teamwork and enables the creation of reproducible snapshots of a project. These snapshots, or commits, capture the exact state of a project at a specific point in time, allowing researchers to reproduce past results or revert to a functioning state if a new update introduces errors. While basic use of a repository is sufficient for transparency (Stage 2), higher levels of reproducibility (Stage 4) require more rigorous versioning, such as "tagging" specific versions of the code to correspond with the results reported in a manuscript or policy brief.

3.4 Try reproducing the research yourself

Reproducibility begins at home. Before sharing findings, ECSs should attempt to reproduce their own work from scratch using only the materials intended for publication. This process serves as a test for the completeness of the documentation and the effectiveness of the organization. For every figure and table included in a manuscript, it is essential to provide the settings, parameters, and tool versions necessary to replicate the results showcased (see Becher et al., 2014, for an example).

Self-replication often highlights pitfalls, such as missing methodological steps or hidden assumptions. A key indicator that documentation needs improvement is if the researcher must rely on their own memory or prior understanding to execute the code; if the instructions are not clear enough for a peer with a similar level of expertise to follow without guidance, the documentation is incomplete. As experience teaches, the minute details of model settings are easily forgotten over time, rendering work irreproducible even to the original

author. Meticulous self-replication following instructions as written confirms that results are a consistent outcome of the methodology described rather than a one-time occurrence.

3.5 Share your work (it's probably good enough)

A common barrier to reproducibility is "perfection paralysis", where researchers hesitate to share findings due to concerns that their work is not yet flawless. However, science thrives on collaboration, feedback, and iterative improvement; embracing this mindset can lead to significant personal growth and scientific advancement. Imperfections are a natural and expected part of research and acknowledging them openly can strengthen a study by identifying alternative approaches or new research opportunities. By sharing data, methods, and code, ECSs contribute to a collective effort that advances knowledge while establishing the reliability and credibility of their own work (Barnes, 2010).

Overcoming the fear of criticism is a significant challenge for ECSs, making it essential to foster a mindset where feedback is viewed as an opportunity for development rather than a professional failure. To mitigate the risk that code or data could be appropriated before a study is fully published, researchers should utilize pre-prints and assign Digital Object Identifiers (DOIs) to their resources. This establishes academic precedence and ownership while facilitating easy citation of reusable "building blocks" such as specific datasets, models, or algorithms (Berger et al., 2024).

A concrete illustration of "good enough" reproducibility is the growing number of modeling studies published on platforms like CoMSES Net, OpenGMS, Github or Zenodo. For example, an agent-based model implementation following the ODD protocol, accompanied by a Jupyter Notebook and archived data, typically meets the "Portable" standard (Stage 3). While such submissions may lack advanced features like automated tests, containerization, or service-based execution (Chen et al., 2020; Moreau et al., 2023) they provide sufficient transparency for the community to engage, identify errors, and build extensions. Ultimately, the benchmark for sharing should not be perfection, but the provision of enough context for independent engagement. Regardless of whether a study involves code, the data and accompanying metadata must be provided in at least human-readable formats (Leipzig et al., 2021). This is a mandatory step for scientific longevity, as the availability of research data declines rapidly with article age, posing a significant challenge to long-term reproducibility (Vines et al., 2014).

To assist researchers in operationalizing these practices during the manuscript preparation phase, we provide a pre-submission checklist (Box 1). This tool is designed to help ECSs verify that their research materials meet at least the recommended Stage 3 (Portable) standard, ensuring that their work is sufficiently transparent and functional for peer review and community engagement.

The practices described above are necessary conditions for improving reproducibility in socio-environmental modelling, yet they are likely insufficient on their own. Each practice places significant demands on the individual researcher – in terms of time, technical skill, and the willingness to share – that are unevenly distributed across career stages. As established in Section 2, ECSs face structural constraints that limit their ability to adopt these practices unilaterally, including institutional mobility, power asymmetries, and the often-marginal nature of reproducibility work relative to primary project obligations. Individual diligence eventually reaches a ceiling beyond which progress depends on shifting the incentive structures, institutional support, and cultural norms that shape research behavior, effectively challenging the current status quo in academia (Micheletti et al., 2024).

4. Moving toward a culture of reproducibility

As established in recent critiques of the academic status quo, the adoption of good modelling practices depends less on the availability of new protocols and more on challenging the systemic barriers of incentives, burden and training (Micheletti et al., 2024). While the scientific community increasingly recognizes the importance of reproducibility, the transition remains slow because the current reward system continues to prioritize publication quantity over methodological rigor. For the ECS, this presents a reproducibility trap: they are the group most likely to establish new norms, yet they have the least power to negotiate the resources required to do so.

Box 1: Pre-submission reproducibility checklist for Early Career Scholars. This checklist identifies the technical requirements for the "Good Enough" baseline (Stage 3: Portable Analysis) necessary for standard peer-reviewed publications, alongside additional markers for robust policy-relevant models (Stage 4) and reusable community software tools (Stage 5).

Stage 3 Baseline: *Achieve this level for most standard peer-reviewed journal publications.*

Documentation and Conceptual Transparency

- A README file exists and describes the purpose of the project, its directory structure, and clear instructions on how to execute the code.
- All data sources are documented, including origin, access date, and a record of any processing or cleaning applied.
- Software dependencies (language version, specific packages, and their versions) are explicitly listed in a configuration file (e.g., requirements.txt or environment.yml).
- The model's conceptual logic and assumptions are described using a standardized protocol (e.g., ODD).
- The rationale for key modelling decisions and methodological pivots is recorded (e.g., in a modelling notebook or as Architectural Decision Records).

Code and Data Organization

- The directory structure is logical, consistently named, and separates raw data from processed data (ensuring raw data is never overwritten).
- All "hard-coded" absolute file paths have been replaced with relative paths or configuration-based path management.
- Code is commented sufficiently for a peer with equivalent expertise to follow the algorithmic logic and execution flow.

Numerical Reproducibility and Self-Verification

- All random seeds have been explicitly fixed and reported for stochastic processes.
- A self-replication test was successful: all figures and tables in the manuscript were successfully regenerated from scratch using only the shared materials.
- The replication was attempted on a clean environment (a different machine or a fresh virtual environment) to ensure portability.

Sharing and Licensing

- The code is stored in a version-controlled repository accessible at a persistent URL (e.g., GitHub, GitLab, or CoMSES Net).
- Data is provided in a human-readable or widely used format with associated metadata.
- An open license (e.g., MIT, GPL, or CC-BY) is explicitly specified for both code and data.

Stage 4: *Include these steps for policy-relevant models or high-stakes management decisions.*

- **TRACE Protocol:** A formal traceability protocol is implemented to document the full modeling cycle.
- **Containerization:** The entire computational environment is encapsulated (e.g., Docker or Singularity).
- **Data Standards:** Data follows formal, machine-readable open standards (e.g., NetCDF).
- **DOI-backed Release:** A DOI (e.g., via Zenodo) has been assigned to the specific version of the materials used in the paper.
- **Numerical Tolerance:** Any known platform-specific numerical differences are documented and within an acceptable tolerance.

Stage 5: *Include these steps for modeling frameworks and software intended for community reuse and extension.*

- **Automated Testing:** The code includes a suite of Unit Tests to verify core logic during future updates.
- **Modular Design:** The code is data-agnostic, modular, and designed for collaborative extension.
- **Advanced Data:** Utilizes cloud-optimized formats (e.g., Zarr or COGS) for large-scale collaboration.
- **Full FAIR Compliance:** The work fully meets all FAIR (Findable, Accessible, Interoperable, Reusable) principles for software.

4.1 Shifting incentives and recognition

While the importance of reproducibility is widely recognized, adoption remains uneven across the scientific community. Although many journals have begun requesting that code and data be made available, such practices are not yet universal nor a mandatory requirement. The adoption by some journals of "code badges" to recognize reproducible workflows (Altmetric, 2016; Call, 2022; Kidwell et al., 2016; Nüst et al., 2018) provides a valuable indicator of changing attitudes. Behind these symbolic rewards is the belief that transparent and well-documented research – particularly work that provides the rationale for decisions taken throughout the modelling process – directly enhances the perceived credibility and trustworthiness of the findings.

Despite these cultural signals, the prevailing measure of success for an ECS remains heavily tied to publication quantity rather than methodological quality. This creates a fundamental conflict between career advancement and the time-intensive work of documenting research. Researchers often hesitate to share code not due to negligence, but due to the significant time required for documentation, the burden of fielding user queries, and the fear that work may be used without proper citation (Stodden, 2010). Furthermore, there are legitimate strategic concerns regarding the loss of competitive advantage and the potential impact on future publications. These are not mere personal preferences; they are rational responses to an academic system that prioritizes the speed and volume of output.

Ultimately, we must convey to ECSs that adopting reproducible practices offers immediate personal benefits that transcend institutional recognition (Micheletti et al., 2024). At the scientific level, it facilitates new collaborations and research opportunities; at the personal level, it makes code easier to test, correct, and repurpose (Allen & Mehler, 2019). These computational skills and the commitment to open science are increasingly recognized as strategic assets on a modern CV (Poldrack, 2019), positioning the ECS as a leader in a shifting scientific landscape.

4.2 Addressing the implementation burden of reproducibility on ECSs

While the importance of reproducibility is established in principle, the practical implementation burden falls disproportionately on ECSs. Even when senior researchers are supportive, it is often the doctoral students and postdocs who perform the time-intensive tasks of cleaning code, documenting data, and managing repositories. This work is frequently treated as a marginal activity rather than a core research obligation, and it is rarely accounted for in project budgets or timelines (Micheletti et al., 2024). Furthermore, ECSs may face skepticism from established academics who argue that the reproducibility crisis is overstated (Fanelli, 2018) or that existing retraction mechanisms are sufficient (Saltelli, 2018; Stodden et al., 2018).

The consequences of this skepticism and the resulting lack of transparency are significant. History illustrates that irreproducibility can lead to high-profile scandals – such as those in cold fusion (Fleischmann & Pons, 1989) or stem cell research (Obokata et al., 2014) – which diminish the achievements of entire fields and erode public trust in science. To ensure scientific integrity and move toward a culture where transparency is the norm (Ioannidis, 2005), institutions and senior researchers must take concrete responsibility for the modeling environment. We offer the following recommendations for each stakeholder group.

For supervisors and research group leaders:

1) Explicitly allocate time within project plans and student milestones for documentation, code review, and data preparation. Reproducibility work that is not scheduled and planned for – such as the transition from a Stage 1 sandbox to a Stage 3 portable workflow – is unlikely to happen.
2) Include reproducibility standards as a named criterion in thesis committee feedback, annual reviews, and project evaluations. These efforts should be explicitly valued alongside traditional output metrics.
3) Participate in the establishment of group-level norms. Rather than delegating the technical setup to ECSs, supervisors should co-develop shared README templates, version control workflows, and data management protocols to ensure consistency across the research group.
4) Model the practices you expect. Senior scientists should share their own code and data, document their modeling decisions, and openly acknowledge the "messiness" of their own earlier work to foster a culture of transparency rather than one of perfectionism.
5) Beta-test the work of Early Career Scientists. Do not leave verification solely to the original author. Before submission or publication, at least one senior co-author should actively attempt to reproduce the results using the documentation and code provided by the student or postdoc. This not only ensures the integrity of the work but also provides a vital mentorship opportunity to debug workflows and improve accessibility.

For institutions and departments:

1) Provide dedicated training integrated into doctoral programs, covering essential computational skills such as version control, data management, and the handling of computational environments.
2) Create, expand, and leverage existing infrastructure that simplifies reproducibility, including shared code repositories, data storage with persistent identifiers, and direct access to research software engineers or data stewards to offload technical hurdles.
3) Recognize reproducibility in hiring and promotion. Code releases, open datasets, and well-documented model repositories should be evaluated as significant scholarly outputs during the recruitment and tenure process.

For funding agencies and journals:

1) Enforce data and code availability as a mandatory condition of funding and publication to ensure that research transparency is not treated as optional.
2) Implement reproducibility checklists at the submission stage (LeVeque et al., 2012; Williams et al., 2018) to help researchers adhere to established guidelines before their work enters the public record.
3) Fund reproducibility activities directly. This could include funding for reproducibility audits, specialized workshops, or “mini sabbaticals” where ECSs can work on documentation and archiving with professional support.

Ultimately, these collective actions are necessary to transform reproducibility from a solitary individual burden into a shared institutional standard. By redistributing the responsibility for implementation, the scientific community can bridge the gap between high-level scientific ideals and the practical constraints of the early-career research stage.

4.3 Standardizing reproducibility processes and bridging the training gaps

Establishing this institutional standard requires a technical baseline that currently does not exist. The research community is producing an exponentially increasing volume of data without a universally codified process for ensuring its reproducibility. Despite efforts to promote concise rules for computational research (Sandve et al., 2013), many ECSs are not formally trained in software development, a discipline that is arguably now integral to socio-environmental modelling. This lack of formal training often necessitates the ad hoc approaches that create long-term barriers to transparency. Furthermore, because the community has yet to converge on a shared definition of what constitutes a minimum level of reproducibility, ECSs are often provided with little guidance on how to maintain these standards throughout the research lifecycle.

The standardization of practices must account for the specific needs of different modeling domains. Research into data science projects indicates that coding violations are frequent and that conventions often need to be context-specific (Simmons et al., 2020). This suggests a need to develop specialized coding standards and automated linting tools tailored to the socio-environmental field to help researchers identify and correct standard violations early in the development process. Because the most pressing global challenges require transdisciplinary integration across the social, physical, and life sciences, education must move beyond theoretical systems thinking toward operationalizing this integration through interoperable workflows (Little et al., 2019; McIntire et al., 2022).

Addressing the reproducibility training gap requires a combination of short-term interventions and long-term curriculum reform. In the short term, professional societies and research institutes should offer dedicated webinars and lecture series on version control, documentation protocols, and collaborative development. However, a sustainable solution depends on integrating reproducibility training into the core undergraduate and graduate curriculum (Kohrs et al., 2023; Micheletti et al., 2024). Introducing these skills as a standard capstone requirement ensures that early-stage researchers acquire the necessary technical proficiency before beginning their professional careers. By fostering a culture where high-quality, reliable research is the baseline expectation, the scientific community can strengthen the overall integrity and public trust in the scientific enterprise (Donoho, 2010). Ensuring reproducibility is a complex task, but through the concerted efforts of researchers, institutions, and publishers, these challenges can be met, leading to a more robust and credible scientific future.

5. A new baseline for reproducibility

Reproducibility is not an all-or-nothing requirement, but a spectrum of technical formalization that must be matched to the purpose and risk of the research. For Early Career Scholars, the goal should not be the attainment of an idealized state of "perfect" reproducibility for every project, but rather the strategic adoption of "good enough" practices that ensure scientific credibility while acknowledging the pragmatic constraints of an early career. By navigating the five stages of computational maturity – from the exploratory sandbox to the reusable tool – ECSs can ensure their contributions are transparent and portable without succumbing to the paralysis often induced by overly rigid standards.

The framework presented here operationalizes the fitness-for-purpose approach (Hamilton et al., 2022) providing a roadmap for individual researchers to determine the appropriate level of investment for their work. Establishing these habits is a strategic career move: it future-proofs research against institutional mobility, enhances scientific impact, and builds the technical proficiency required for modern, interdisciplinary socio-environmental modelling. However, as we have argued, individual diligence eventually reaches a ceiling. A sustainable culture of reproducibility requires a fundamental shift in how institutions, supervisors, and funding agencies value and resource the technical labor behind reproducible modelling (Micheletti et al., 2024).

Ultimately, the responsibility for scientific integrity is a shared one. By adopting "good enough" practices today, ECSs are not merely navigating a crisis – they are establishing a new professional baseline. As socio-environmental challenges grow in complexity and high-stakes policy decisions depend increasingly on model outputs, the demand for transparent and auditable workflows will only intensify. Challenging the academic status quo and prioritizing reproducibility is the first step toward a more robust, collaborative, and trustworthy scientific future.

Acknowledgements

We thank Volker Grimm for comments on earlier drafts and Samuel Fischer for early discussions during the drafting process. This work also benefited from the organizing efforts and community structure of the Open Modeling Foundation and discussions within the OMF Early Career Scholars Working Group. RJ was funded by MSTDI RS (no. 451-03-33/2026-03/200009 from 5.2.2026) and ADAPTATUR (PID2020114186RB-C22, Ministry of Science and Innovation, Government of Spain); TM by the Deutsche Forschungsgemeinschaft (DFG, project 529743012) and BirdFutures (BMFTR, project 033L313A-B); and JS by the TU Delft AI Labs and Talent Program during the development of this manuscript.

References

Allen, C., Mehler, D.M.A., 2019. Open science challenges, benefits and tips in early career and beyond. PLOS Biol. 17, e3000246. https://doi.org/10.1371/journal.pbio.3000246

Altmetric, 2016. Free Badges for Researchers. Altmetric. URL https://www.altmetric.com/products/free-tools/free-badges-for-researchers/ (accessed 2.26.23).

Ayllón, D., Railsback, S.F., Gallagher, C., Augusiak, J., Baveco, H., Berger, U., Charles, S., Martin, R., Focks, A., Galic, N., Liu, C., van Loon, E.E., Nabe-Nielsen, J., Piou, C., Polhill, J.G., Preuss, T.G., Radchuk, V., Schmolke, A., Stadnicka-Michalak, J., Thorbek, P., Grimm, V., 2021. Keeping modelling notebooks with TRACE: Good for you and good for environmental research and management support. Environ. Model. Softw. 136, 104932. https://doi.org/10.1016/j.envsoft.2020.104932

Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454. https://doi.org/10.1038/533452a

Barciauskas, A., Mandel, A., Barron, K., Deziel, Z., 2023. Cloud-Optimized Geospatial Formats Guide [WWW Document]. Cloud-Optim. Geospatial Formats Guide. URL https://guide.cloudnativegeo.org/ (accessed 5.12.24).

Barnes, N., 2010. Publish your computer code: it is good enough. Nature 467, 753–753. https://doi.org/10.1038/467753a

Barton, C.M., Ames, D., Chen, M., Frank, K., Jagers, H.R.A., Lee, A., Reis, S., Swantek, L., 2022a. Making modeling and software FAIR. Environ. Model. Softw. 156, 105496. https://doi.org/10.1016/j.envsoft.2022.105496

Barton, C.M., Lee, A., Janssen, M.A., van der Leeuw, S., Tucker, G.E., Porter, C., Greenberg, J., Swantek, L., Frank, K., Chen, M., Jagers, H.R.A., 2022b. How to make models more useful. Proc. Natl. Acad. Sci. U.S.A. 119, e2202112119. https://doi.org/10.1073/pnas.2202112119

Becher, M.A., Grimm, V., Thorbek, P., Horn, J., Kennedy, P.J., Osborne, J.L., 2014. BEEHAVE: a systems model of honeybee colony dynamics and foraging to explore multifactorial causes of colony failure. J. Appl. Ecol. 51, 470–482.

https://doi.org/10.1111/1365-2664.12222

Benz, J., Hoch, R., Legović, T., 2001. ECOBAS — modelling and documentation. Ecol. Model. 138, 3–15. https://doi.org/10.1016/S0304-3800(00)00389-6

Berger, U., Bell, A., Barton, C.M., Chappin, E., Dreßler, G., Filatova, T., Fronville, T., Lee, A., van Loon, E., Lorscheid, I., Meyer, M., Müller, B., Piou, C., Radchuk, V., Roxburgh, N., Schüler, L., Troost, C., Wijermans, N., Williams, T.G., Wimmler, M.-C., Grimm, V., 2024. Towards reusable building blocks for agent-based modelling and theory development. Environ. Model. Softw. 175, 106003. https://doi.org/10.1016/j.envsoft.2024.106003

Bhandari Neupane, J., Neupane, R.P., Luo, Y., Yoshida, W.Y., Sun, R., Williams, P.G., 2019. Characterization of Leptazolines A–D, Polar Oxazolines from the Cyanobacterium Leptolyngbya sp., Reveals a Glitch with the "Willoughby–Hoye" Scripts for Calculating NMR Chemical Shifts. Org. Lett. 21, 8449–8453. https://doi.org/10.1021/acs.orglett.9b03216

Brinckman, A., Chard, K., Gaffney, N., Hategan, M., Jones, M.B., Kowalik, K., Kulasekaran, S., Ludäscher, B., Mecum, B.D., Nabrzyski, J., Stodden, V., Taylor, I.J., Turk, M.J., Turner, K., 2019. Computing environments for reproducibility: Capturing the "Whole Tale." Future Generation Computer Systems 94, 854–867. https://doi.org/10.1016/j.future.2017.12.029

Buckheit, J.B., Donoho, D.L., 1995. WaveLab and Reproducible Research, in: Antoniadis, A., Oppenheim, G. (Eds.), Wavelets and Statistics, Lecture Notes in Statistics. Springer New York, New York, NY, pp. 55–81. https://doi.org/10.1007/978-1-4612-2544-7_5

Call, M., 2022. Badges for Open Research Practices Now Available on OSF Registrations [WWW Document]. URL https://www.cos.io/blog/badges-for-open-research-practices-available-on-osf-registrations (accessed 7.23.23).

Chen, M., Voinov, A., Ames, D.P., Kettner, A.J., Goodall, J.L., Jakeman, A.J., Barton, M.C., Harpham, Q., Cuddy, S.M., DeLuca, C., Yue, S., Wang, J., Zhang, F., Wen, Y., Lü, G., 2020. Position paper: Open web-distributed integrated geographic modelling and simulation to enable broader participation and applications. Earth-Sci. Rev. 207, 103223. https://doi.org/10.1016/j.earscirev.2020.103223

Committee on Reproducibility and Replicability in Science, Board on Behavioral, Cognitive, and Sensory Sciences, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Nuclear and Radiation Studies Board, Division on Earth and Life Studies, Board on Mathematical Sciences and Analytics, Committee on Applied and Theoretical Statistics, Division on Engineering and Physical Sciences, Board on Research Data and Information, Committee on Science, Engineering, Medicine, and Public Policy, Policy and Global Affairs, National Academies of Sciences, Engineering, and Medicine, 2019. Reproducibility and Replicability in Science. National Academies Press, Washington, D.C. https://doi.org/10.17226/25303

Community Surface Dynamics Modeling System (CSDMS), 2026. Awards: Student Modeler Award.

Data Carpentry Committee, 2016. Introduction to Reproducible Research with `R`. Data Carpentry Committee. https://github.com/data-carpentry/rr-intro

Donoho, D., 2024. Data Science at the Singularity. Harv. Data Sci. Rev. 6.1. https://doi.org/10.1162/99608f92.b91339ef

Donoho, D.L., 2010. An invitation to reproducible computational research. Biostatistics 11, 385–388. https://doi.org/10.1093/biostatistics/kxq028

Edmonds, B., 2016. Review of The Aqua Book: Guidance on Producing Quality Analysis for Government [WWW Document]. URL https://jasss.soc.surrey.ac.uk/19/3/reviews/7.html (accessed 6.1.26).

Enserink, M., 2021. Does ocean acidification alter fish behavior? Fraud allegations create a sea of doubt [WWW Document]. Science.org. URL https://www.science.org/content/article/does-ocean-acidification-alter-fish-behavior-fraud-allegations-create-sea-doubt (accessed 1.21.24).

Esbaugh, A.J., 2023. Recalibrating the significance of the decline effect in fish ocean acidification research. PLOS Biol. 21, e3002113. https://doi.org/10.1371/journal.pbio.3002113

Fanelli, D., 2018. Is science really facing a reproducibility crisis, and do we need it to? Proc. Natl. Acad. Sci. 115, 2628–2631. https://doi.org/10.1073/pnas.1708272114

Filatova, T., Verbeek, L., Warnier, M., Ghorbani, A., Nikolic, I., Grimm, V., Berger, U., Barton, M., Bell, A., Lee, A., Magliocca, N.R., Wagenblast, T., 2025. AGENTBLOCKS: A Community Platform for Sharing, Comparing, and Improving Reusable Building Blocks for (Agent-Based) Models. J. Artif. Soc. Soc. Simul. 28, 11. https://doi.org/10.18564/jasss.5831

Fleischmann, M., Pons, S., 1989. Electrochemically induced nuclear fusion of deuterium. J. Electroanal. Chem. Interfacial Electrochem. 261, 301–308. https://doi.org/10.1016/0022-0728(89)80006-3

Grimm, V., 2023. Ecology needs to overcome siloed modelling. Trends Ecol. Evol. 38, 1122–1124. https://doi.org/10.1016/j.tree.2023.09.011

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Vabø, R., Visser, U., DeAngelis, D.L., 2006. A standard protocol for describing individual-based and agent-based models. Ecol. Model. 198, 115–126. https://doi.org/10.1016/j.ecolmodel.2006.04.023

Grimm, V., Railsback, S.F., Vincenot, C.E., Berger, U., Gallagher, C., DeAngelis, D.L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A.S.A., Milles, A., Nabe-Nielsen, J., Polhill, J.G., Radchuk, V., Rohwäder, M.-S., Stillman, R.A., Thiele, J.C., Ayllón, D., 2020. The ODD protocol for describing agent-based and other simulation models: a second update to improve clarity, replication, and structural realism. J. Artif. Soc. Soc. Simul. 23, 7. https://doi.org/10.18564/jasss.4259

Hamilton, S.H., Pollino, C.A., Stratford, D.S., Fu, B., Jakeman, A.J., 2022. Fit-for-purpose environmental modeling: Targeting the intersection of usability, reliability and feasibility. Environ. Model. Softw. 148, 105278.

https://doi.org/10.1016/j.envsoft.2021.105278
Higginson, A.D., Munafò, M.R., 2016. Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. PLOS Biol. 14, e2000995. https://doi.org/10.1371/journal.pbio.2000995
HM Treasury, 2015. Aqua Book: Guidance on producing quality analysis for government.
Hong, N.P.C., Crick, T., Gent, I.P., Kotthoff, L., Takeda, K., 2015. Top Tips to Make Your Research Irreproducible. https://doi.org/10.48550/arXiv.1504.00062
Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., Arheimer, B., 2016. Most computational hydrology is not reproducible, so is it really science? Water Resour. Res. 52, 7548–7555. https://doi.org/10.1002/2016WR019285
Hutton, E.W.H., Piper, M.D., Tucker, G.E., 2020. The Basic Model Interface 2.0: A standard interface for coupling numerical models in the geosciences. J. Open Source Softw. 5, 2317. https://doi.org/10.21105/joss.02317
Huxley, A., Jones, S., Bradshaw, D., 1994. The human situation: lectures at Santa Barbara, 1959, Flamingo modern classic. Flamingo, London.
Ioannidis, J.P.A., 2005. Why Most Published Research Findings Are False. PLOS Med. 2, 2–8. https://doi.org/10.1371/journal.pmed.0020124
Iwanaga, T., Wang, H.-H., Hamilton, S.H., Grimm, V., Koralewski, T.E., Salado, A., Elsawah, S., Razavi, S., Yang, J., Glynn, P., Badham, J., Voinov, A., Chen, M., Grant, W.E., Peterson, T.R., Frank, K., Shenk, G., Barton, C.M., Jakeman, A.J., Little, J.C., 2021a. Socio-technical scales in socio-environmental modeling: Managing a system-of-systems modeling approach. Environ. Model. Softw. 135, 104885. https://doi.org/10.1016/j.envsoft.2020.104885
Iwanaga, T., Wang, H.-H., Koralewski, T.E., Grant, W.E., Jakeman, A.J., Little, J.C., 2021b. Toward a complete interdisciplinary treatment of scale: Reflexive lessons from socioenvironmental systems modeling. Elem. Sci. Anthr. 9. https://doi.org/10.1525/elementa.2020.00182
James, S.R., Foks, N.L., Minsley, B.J., 2022. GSPy: A new toolbox and data standard for Geophysical Datasets. Front. Earth Sci. 10. https://doi.org/10.3389/feart.2022.907614
Jebeile, J., Crucifix, M., 2020. Multi-model ensembles in climate science: Mathematical structures and expert judgements. Stud. Hist. Philos. Sci. Part A 83, 44–52. https://doi.org/10.1016/j.shpsa.2020.03.001
Jenkins, G.B., Beckerman, A.P., Bellard, C., Benítez-López, A., Ellison, A.M., Foote, C.G., Hufton, A.L., Lashley, M.A., Lortie, C.J., Ma, Z., Moore, A.J., Narum, S.R., Nilsson, J., O'Boyle, B., Provete, D.B., Razgour, O., Rieseberg, L., Riginos, C., Santini, L., Sibbett, B., Peres-Neto, P.R., 2023. Reproducibility in ecology and evolution: Minimum standards for data and code. Ecol. Evol. 13, e9961. https://doi.org/10.1002/ece3.9961
Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C.L., Errington, T., Timothy M Errington, Fiedler, S., Brian A. Nosek, Nosek, B.A., Brian A. Nosek, 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLoS Biol. 14(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456
Kim, Y.-M., Poline, J.-B., Dumas, G., 2018. Experimenting with reproducibility: a case study of robustness in bioinformatics. GigaScience 7, giy077. https://doi.org/10.1093/gigascience/giy077
Kohrs, F.E., Auer, S., Bannach-Brown, A., Fiedler, S., Haven, T.L., Heise, V., Holman, C., Azevedo, F., Bernard, R., Bleier, A., Bössel, N., Cahill, B.P., Castro, L.J., Ehrenhofer, A., Eichel, K., Frank, M., Frick, C., Friese, M., Gärtner, A., Gierend, K., Grüning, D.J., Hahn, L., Hülsemann, M., Ihle, M., Illius, S., König, L., König, M., Kulke, L., Kutlin, A., Lammers, F., Mehler, D.M., Miehl, C., Müller-Alcazar, A., Neuendorf, C., Niemeyer, H., Pargent, F., Peikert, A., Pfeuffer, C.U., Reinecke, R., Röer, J.P., Rohmann, J.L., Sánchez-Tójar, A., Scherbaum, S., Sixtus, E., Spitzer, L., Straßburger, V.M., Weber, M., Whitmire, C.J., Zerna, J., Zorbek, D., Zumstein, P., Weissgerber, T.L., 2023. Eleven strategies for making reproducible research and open science training the norm at research institutions. eLife 12, e89736. https://doi.org/10.7554/eLife.89736
Kondyli, D., Nisiotis, C.-S., Klironomos, N., 2024. Data reusability for migration research: a use case from SoDaNet data repository. Front. Hum. Dyn. 5, 1310420. https://doi.org/10.3389/fhumd.2023.1310420
Kruchten, P., 2008. What do software architects really do? J. Syst. Softw. 81, 2413–2416. https://doi.org/10.1016/j.jss.2008.08.025
Lagisz, M., Aich, U., Amin, B., Rutkowska, J., Sánchez-Mercado, A., Lara, C.E., Nakagawa, S., 2023. Little transparency and equity in scientific awards for early- and mid-career researchers in ecology and evolution. Nat. Ecol. Evol. 7, 655–665. https://doi.org/10.1038/s41559-023-02028-6
Leek, J.T., Jager, L., 2016. Is most published research really false. bioRxiv 050575. https://doi.org/10.1101/050575
Leipzig, J., Nüst, D., Hoyt, C.T., Ram, K., Greenberg, J., 2021. The role of metadata in reproducible computational research. 2, 100322. https://doi.org/10.1016/j.patter.2021.100322
LeVeque, R.J., Mitchell, I.M., Stodden, V., 2012. Reproducible research for scientific computing: Tools and strategies for changing the culture. Comput. Sci. Eng. 14, 13–17. https://doi.org/10.1109/mcse.2012.38
Little, J.C., Hester, E.T., Elsawah, S., Filz, G.M., Sandu, A., Carey, C.C., Iwanaga, T., Jakeman, A.J., 2019. A tiered, system-of-systems modeling framework for resolving complex socio-environmental policy issues. Environ. Model. Softw. 112, 82–94. https://doi.org/10.1016/j.envsoft.2018.11.011
McConnell, S., 2004. Code Complete, Second Edition, 2nd ed. Pearson Education.
McIntire, E.J.B., Chubaty, A.M., Cumming, S.G., Andison, D., Barros, C., Boisvenue, C., Haché, S., Luo, Y., Micheletti, T., Stewart, F.E.C., 2022. PERFICT: A Re-imagined foundation for predictive ecology. Ecol. Lett. 25, 1345–1351. https://doi.org/10.1111/ele.13994
Meirmans, S., Butlin, R.K., Charmantier, A., Engelstädter, J., Groot, A.T., King, K.C., Kokko, H., Reid, J.M., Neiman, M., 2019.

Science policies: How should science funding be allocated? An evolutionary biologists' perspective. J. Evol. Biol. 32, 754–768. https://doi.org/10.1111/jeb.13497

Melsen, L.A., Teuling, A.J., Torfs, P.J.J.F., Zappa, M., Mizukami, N., Mendoza, P.A., Clark, M.P., Uijlenhoet, R., 2019. Subjective modeling decisions can significantly impact the simulation of flood and drought events. J. Hydrol. 568, 1093–1104. https://doi.org/10.1016/j.jhydrol.2018.11.046

Melsen, L.A., Torfs, P.J.J.F., Uijlenhoet, R., Teuling, A.J., 2017. Comment on "Most computational hydrology is not reproducible, so is it really science?" by Christopher Hutton et al. Water Resour. Res. 53, 2568–2569. https://doi.org/10.1002/2016WR020208

Micheletti, T., Wimmler, M.-C., Berger, U., Grimm, V., McIntire, E.J., 2024. Beyond guides, protocols and acronyms: Adoption of good modelling practices depends on challenging academia's status quo in ecology. Ecol. Model. 496, 110829. https://doi.org/10.1016/j.ecolmodel.2024.110829

Moreau, D., Wiebels, K., Boettiger, C., 2023. Containers for computational reproducibility. Nat. Rev. Methods Primer 3, 50. https://doi.org/10.1038/s43586-023-00236-9

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F.O., Sileryte, R., Cerutti, V., 2018. Reproducible research and GIScience: an evaluation using AGILE conference papers. PeerJ 6, 1–23. https://doi.org/10.7717/peerj.5072

Obokata, H., Wakayama, T., Sasai, Y., Kojima, K., Vacanti, M.P., Niwa, H., Yamato, M., Vacanti, C.A., 2014. Retraction Note: Stimulus-triggered fate conversion of somatic cells into pluripotency. Nature 511, 112–112. https://doi.org/10.1038/nature13598

Open Science Collaboration. Estimating the reproducibility of psychological science. Science 349, 943–943. https://doi.org/10.1126/science.aac4716

Pellizzari, E., Lohr, K., Blatecky, A., Creel, D., 2017. Reproducibility: A Primer on Semantics and Implications for Research. RTI Press. https://doi.org/10.3768/rtipress.2017.bk.0020.1708

Plesser, H.E., 2018. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. Front. Neuroinformatics 11, 76–76. https://doi.org/10.3389/fninf.2017.00076

Poldrack, R.A., 2019. The Costs of Reproducibility. Neuron 101, 11–14. https://doi.org/10.1016/j.neuron.2018.11.030

Popper, K.R., 1959. The logic of scientific discovery, 2008 reprint. ed. Routledge, London ; New York.

Rekha V, S., Muccini, H., 2018. Group decision-making in software architecture: A study on industrial practices. Inf. Softw. Technol. 101, 51–63. https://doi.org/10.1016/j.infsof.2018.04.009

Saltelli, A., 2018. Why science's crisis should not become a political battling ground. Futures 104, 85–90. https://doi.org/10.1016/j.futures.2018.07.006

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why So Many Published Sensitivity Analyses Are False: A Systematic Review of Sensitivity Analysis Practices. Environ. Model. Softw. 114, 29–39.

Saltelli, A., Funtowicz, S., 2017. What is science's crisis really about? Futures, Post-Normal science in practice 91, 5–11. https://doi.org/10.1016/j.futures.2017.05.010

Sandve, G.K., Nekrutenko, A., Taylor, R.J., Hovig, E., 2013. Ten simple rules for reproducible computational research. PLOS Comput. Biol. 9(10). https://doi.org/10.1371/journal.pcbi.1003285

Schnell, S., 2018. "Reproducible" Research in Mathematical Sciences Requires Changes in our Peer Review Culture and Modernization of our Current Publication Approach. Bull. Math. Biol. 80, 3095–3105. https://doi.org/10.1007/s11538-018-0500-9

Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D., Kuemmerlen, M., Langhans, S.D., Martínez-López, J., Vermeiren, P., 2019. How to make ecological models useful for environmental management. Ecol. Model. 411, 108784. https://doi.org/10.1016/j.ecolmodel.2019.108784

Seppelt, R., Richter, O., 2005. "It was an artefact not the result": A note on systems dynamic model development tools. Environ. Model. Softw., Environmental Knowledge and Information Systems 20, 1543–1548. https://doi.org/10.1016/j.envsoft.2004.12.004

Samuel, S., König-Ries, B., 2021. Understanding experiments and research practices for reproducibility: an exploratory study. PeerJ 9, e11140. https://doi.org/10.7717/peerj.11140

Simmons, A.J., Barnett, S., Rivera-Villicana, J., Bajaj, A., Vasa, R., 2020. A large-scale comparative analysis of Coding Standard conformance in Open-Source Data Science projects, in: Proceedings of the 14th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM). Presented at the ESEM '20: ACM / IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, Bari Italy, pp. 1–11. https://doi.org/10.1145/3382494.3410680

Sletholt, M.T., Hannay, J.E., Pfahl, D., Langtangen, H.P., 2012. What do we know about scientific software development's agile practices? Computing in Science & Engineering 14(2), 24–36. https://doi.org/10.1109/MCSE.2011.113

Stodden, V., 2010. The Scientific Method in Practice: Reproducibility in the Computational Sciences. https://doi.org/10.2139/ssrn.1550193

Stodden, V., Seiler, J., Ma, Z., 2018. An empirical analysis of journal policy effectiveness for computational reproducibility. Proc. Natl. Acad. Sci. U. S. A. 115, 2584–2589. https://doi.org/10.1073/pnas.1708290115

Theiler, J., 1990. Estimating Fractal Dimension. J Opt Soc Am A 7, 1055–1073.

Vines, T.H., Albert, A., Andrew, R.L., Débarre, F., Bock, D.G., Franklin, M.T., Gilbert, K.J., Prunier, J.G., Moore, J.-S., Renaut, S., Rennison, D.J., 2014. The Availability of Research Data Declines Rapidly with Article Age. Curr. Biol. 24, 94–97. https://doi.org/10.1016/j.cub.2013.11.014

Wang, T., Wang, S., Chen, T.-H. (Peter), 2023. Study the correlation between the readme file of GitHub projects and their popularity. J. Syst. Softw. 205, 111806. https://doi.org/10.1016/j.jss.2023.111806

Williams, M., Mullane, K., Curtis, M.J., 2018. Addressing Reproducibility: Peer Review, Impact Factors, Checklists, Guidelines, and Reproducibility Initiatives, In: Research in the Biomedical Sciences. Elsevier, pp. 197–306. https://doi.org/10.1016/b978-0-12-804725-5.00005-7

Wilson, G., Aruliah, D.A., Brown, C.T., Hong, N.P.C., Davis, M., Guy, R.T., Haddock, S.H.D., Huff, K.D., Mitchell, I.M., Plumbley, M.D., Waugh, B., White, E.P., Wilson, P., 2014. Best Practices for Scientific Computing. PLOS Biol. 12, e1001745. https://doi.org/10.1371/journal.pbio.1001745

Wilson, J.P., Butler, K., Gao, S., Hu, Y., Li, W., Wright, D.J., 2021. A Five-Star Guide for Achieving Replicability and Reproducibility When Working with GIS Software and Algorithms. Ann. Am. Assoc. Geogr. 111(5), 1311-1317. https://doi.org/10.1080/24694452.2020.1806026

Wirtz, D., Nowak, W., 2017. The rocky road to extended simulation frameworks covering uncertainty, inversion, optimization and control. Environ. Model. Softw. 93, 180–192. https://doi.org/10.1016/j.envsoft.2016.10.003

Zhu, Z., Chen, M., Sun, L., Qian, Z., He, Y., Ma, Z., Zhang, F., Wen, Y., Yue, S., Lü, G., 2023. Reproducing computational processes in service-based geo-simulation experiments. Int. J. Appl. Earth Obs. Geoinformation 124, 103520. https://doi.org/10.1016/j.jag.2023