# PNAS

💬 OPINION

How to make models more useful

C. Michael Barton[a,b,c,1] , Allen Lee[a], Marco A. Janssen[a,d] , Sander van der Leeuw[e] , Gregory E. Tucker[f,g] ,
Cheryl Porter[h] , Joshua Greenberg[i] , Laura Swantek[a] , Karin Frank[j] , Min Chen[k,l] , and H. R. Albert Jagers[m]
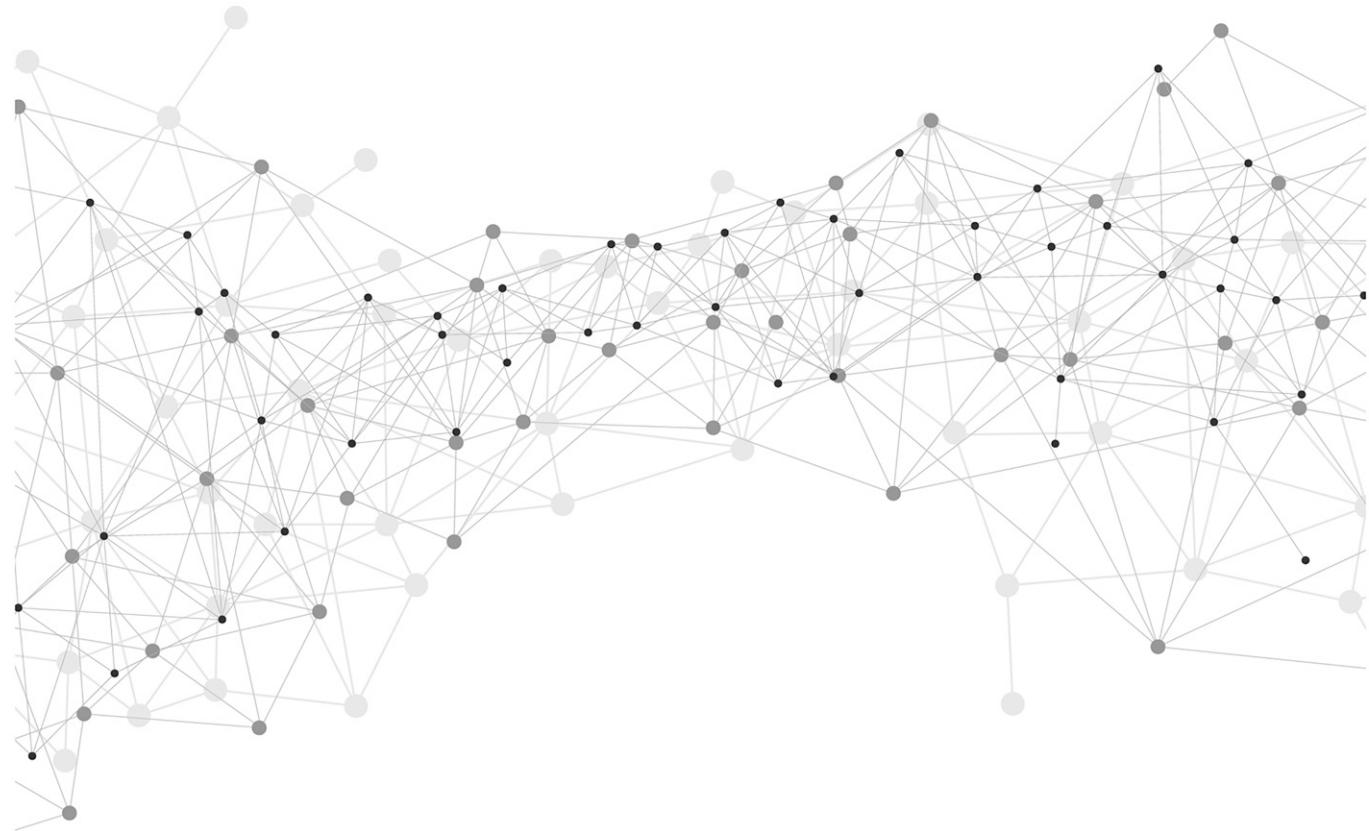
Computational modeling has become a valuable tool for science and policy, but community standards to share model details have not kept pace. For research to be replicated, evaluated, and improved, it's important that model code—written in a way that is comprehensible, transparent, and, in an ideal world, easily executable—be preserved alongside the published articles that describe the results. This is not yet the case for most modeling science. To respond to this challenge, the scientific modeling community has come together to promote best-practice standards for publishing model code and support professional incentives to encourage their adoption. But we need to do more.

In a wide range of research fields, computational modeling has become a critical tool. Its use has grown to augment and even replace narrative and mathematical representations of societal and biophysical processes. Models allow researchers to represent and study complex, dynamic interactions of multiple processes in ways not possible with more traditional means. Applications of computational modeling span the evolution of galaxies to subatomic physics, Earth tectonics, global temperature change, sea-level rise, emergence and loss of biodiversity, economies, crop yields, and the spread of misinformation—not to mention the ongoing reverberations of a global pandemic.

But there's a problem: Articles that report the results of models are frequently not sufficient to reproduce the models, even when the articles describe the underlying concepts and assumptions. The source code of the models—the human-readable program created by its programmer—must also be accessible, understandable, and runnable by others. This is especially important when computational models become a primary laboratory for scientific research and the

Computational models are increasingly valuable tools. But articles that report the results of models are frequently not sufficient to reproduce the models. We need to do a better job making the source code of models accessible, understandable, and runnable by others. Image credit: Image credit: Shutterstock/NWM.

basis for high-impact policy decisions regarding such things as climate change and disease spread.

Models and Open Science

The past decade has seen a movement to promote open sharing of data and software that underlie scientific research by adopting a set of "FAIR" principles [Findability, Accessibility, Interoperability, and Reusability (1)]. Yet, a recent study of nearly 8,000 articles on model-based research from 1990 through 2018, listed in ISI Web of Science, found that a majority do not make the model code available (Fig. 1) (2). Even for the most recent articles in the study, more than 80% do not provide access to the model code. Researchers share the results of model-based research in peer-reviewed journals, following widely understood and accepted scientific norms. But there are no equivalent formal or informal standards within the scientific community for how model code should be made available, which model version should be shared, how the code should be documented, or how it should be packaged—so that it can be run effectively, be compiled if needed, or coupled with other models to represent interacting social and natural processes. A growing number of journals now recommend or require that authors make available data on which research is based. Funding agencies, including the US National Science Foundation, National Institutes of Health, and Department of Agriculture, now require data management plans, as do EU science funding agencies. Some steps have been taken to adapt FAIR principles to research software (3–5) but the community, including editors and funders, still have no guidelines on how to apply them to model code.

This lack of community-wide standards impedes scientific innovation. A researcher inspired by new model-based research, but lacking access to the documented and runnable code, must reverse-engineer algorithms from the usually inadequate description in a journal article. Others try to adapt available (but often inappropriate) models, outsource modeling to someone unfamiliar with the relevant science questions, or simply give up in frustration. Those who do manage to reconstruct models from published research proceed more slowly and can repeat errors and coding inefficiencies that could be avoided with access to understandable and runnable code. And requesting the source code from model creators, although it might sound like a straightforward remedy, often fails to satisfy. Either the authors don't respond or, if they do, they have difficulty providing an understandable, runnable version of model code that was used to generate published research (6).

There is another problem. Without access to the original code, the results of model-based science cannot be fully evaluated in peer review, which has repercussions for the "reproducibility crisis" in science and erodes public trust (7). An important reason that the 2009 so-called "climategate" affair (resulting from the hacking and release of emails by climate researchers at the University of East Anglia in the UK) had such a negative impact on public confidence in climate science was the lack of scientific transparency, including restricted access to climate models and data sets (8, 9). This is somewhat ironic because climate models have some of the most rigorously tested and reliable scientific code (10). Yet more than a decade later, little has changed. The Community Earth System Model (CESM, supported by the US National Center for Atmospheric Research) is still one of the few climate models used in the recent reports by the Intergovernmental Panel on Climate Change (IPCC) to make its code and data openly accessible and documented. Likewise, the lack of transparency and open access to understandable and runnable code for epidemiological models reported by researchers and used for public health policy during the initial months of the coronavirus pandemic contributed to politicization, polarization, and confusion among governments and the general populace (11).

Why Models Are Not FAIR

The open, accessible, and extensively documented CESM, mentioned above, is complicated to download and difficult to install and run. Even so, it has been downloaded 26,590 times from 2015 to 2020 (12), and its use has generated
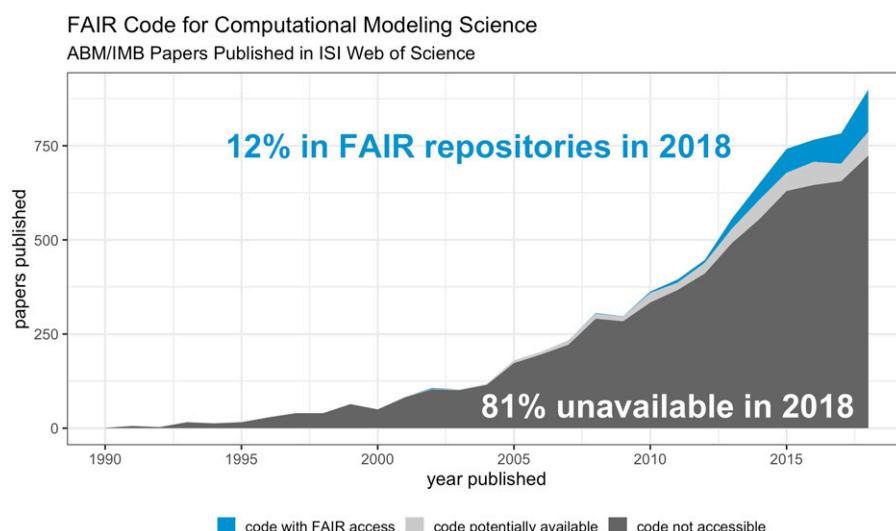

Fig. 1. Articles presenting results of agent-based and individual-based models from 1990–2018. Code with FAIR access refers to code published in persistent, trusted, FAIR-aligned repositories. Code not accessible refers to articles in which the authors do not indicate any location from which code can be downloaded. Image credit: Data from Ref (2) and used with permission.

nearly 500 data sets openly published on Zenodo, Open Science Framework, and GitHub. The Community Surface Dynamics Modeling System (CSDMS) provides a FAIR-aligned repository for computational models of geophysical systems that have been cited over 720,000 times (13). In the FAIR-aligned model library for social and ecological models at CoMSES.Net (Network for Computational Modeling in Social and Ecological Sciences), archived codes have been downloaded more than 230,000 times in the past five years. The growth of the CSDMS and CoMSES.Net repositories shows that an increasing number of researchers recognize the value of FAIR-aligned code (Fig. 2).

Why, then, is so little model code discoverable, accessible, reusable, or interoperable? Many of the causes articulated a decade ago still persist (14). Although IPR issues and other institutional restrictions on code dissemination are common for commercial software, they rarely apply to scientific models. Instead, the obstacles fall into two broad, interrelated categories: Model developers are concerned they will not receive professional recognition or rewards for their work, and they say they cannot afford the time and effort needed to ensure their code is good enough to publish in alignment with FAIR principles (15).

Modelers worry that if they make their code openly accessible others may use it without giving them credit or, worse, falsely claim authorship. There are concerns that if model code is used and cited in another work, the original developer may have no way to document this, or that their employer might not classify the effort as contributing to professional advancement (16). Also, computational models are often produced quickly with evolving, exploratory goals and strict resource constraints (17). Making model code accessible, findable, citable in a trusted digital repository, and transparent and documented enough that it can be compiled, executed, and understood requires significant additional work. These issues are especially acute for early-career researchers.

Addressing these problems and concerns involves greater awareness of open source licensing, the introduction of widespread ethical norms to cite research software, ensuring citations of FAIR models appear in citation indices, and recognition and rewards from professional bodies for research that complies with the FAIR code. In other words, we need institutional incentives for FAIR practice in modeling, broadly supported by the modeling community, scientific publishers, funders, and academic and research organizations that employ modeling researchers.

International Collaboration

To meet these challenges, representatives of leading organizations that support computational modeling across the social, biomedical, ecological, environmental, and geophysical sciences met in December 2021 to establish the Open Modeling Foundation (OMF). A central mission of the OMF is to adopt existing standards or develop new ones, if needed, to help modeling researchers, research and academic organizations, journals, funders, and other stakeholders to define what it means for a model to be FAIR. It will also offer guidance to help researchers meet these standards.

Although such standards are needed, they are not sufficient. Equally important are incentives to recognize and reward researchers who undertake the additional work needed to make their code FAIR: preparing clean and commented code for archiving in a trusted digital repository, writing documentation to make a model reusable, submitting code for peer review, and possibly packaging code with all the supporting software it needs to run and an application programming interface (API) that will allow it to be coupled to other models.

The OMF will help establish these incentives by:

- recommending official digital badges or other public markers for articles or digital repositories that recognize FAIR code,
- establishing guidelines for evaluating digitally published code,
- adopting common standards for code citation and metadata in publications,
- serving as a clearinghouse for educational materials on implementing FAIR practices,
- promoting the ethics of properly citing FAIR code with the many thousands of modeling researchers represented by OMF member organizations, and
- influencing employers of modeling researchers to recognize the value of publishing FAIR model code for professional advancement.

Fortunately, the OMF can build on existing projects to advance these goals. For example, OMF member the Research Data Alliance US has collaborated with Force11
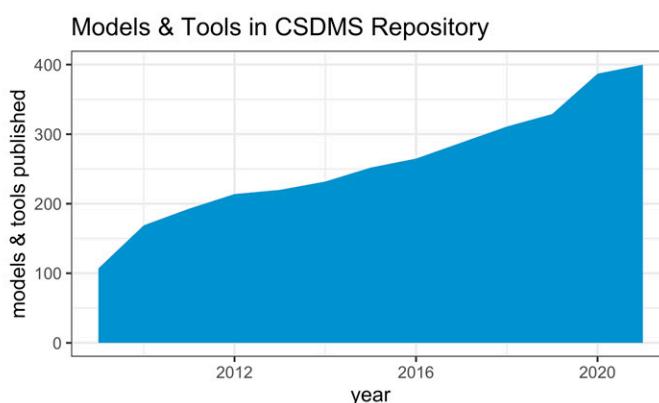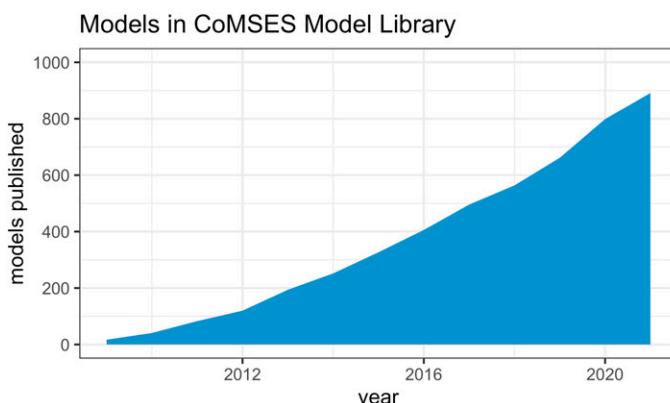

Fig. 2. History of model code published in FAIR-aligned repositories of CoMSES.Net and CSDMS. Image credit: Data compiled by Barton and Lee for CoMSES.Net and Tucker for CSDMS.

and the Research Software Alliance to draft FAIR standards for scientific software (5), which can be adapted for models and adopted across the scientific modeling community. Other OMF member efforts include improving citation and metadata standards for digital data and code by the American Geophysical Union, protocols for documenting individual and agent-based models published by the Helmholtz-Centre for Environmental Research Ltd (18), and standard APIs for model interoperability under development by CSDMS and the Key Laboratory of Virtual Geographic Environment at Nanjing Normal University (19, 20).

What else can the OMF offer beyond these initiatives? By acting as a central hub for these and other FAIR standards for computational modeling, these important ideas can be more broadly discussed, enhanced, disseminated, and adopted.

We call on all researchers, modeling science organizations, and other stakeholders to support the OMF and its goals of making scientific models FAIR and rewarding those who do so. We invite any formally constituted organization that supports or represents modeling science to join the Open Modeling Foundation and also invite individual modeling researchers to join OMF Working Groups (see https://openmodelingfoundation.github.io).

Author affiliations: [a]School of Complex Adaptive Systems, Arizona State University, Tempe, AZ 85287; [b]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287; [c]Network for Computational Modeling in Social and Ecological Sciences, Arizona State University, Tempe, AZ 85287; [d]School of Sustainability, Arizona State University, Tempe, AZ 85287; [e]Santa Fe Institute, Santa Fe, NM 87501; [f]Geological Sciences, University of Colorado Boulder, Boulder, CO 80309; [g]Community Surface Dynamics Modeling System, University of Colorado Boulder, Boulder, CO 80309; [h]Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611; [i]Alfred P. Sloan Foundation, New York, NY 10111; [j]Department of Ecological Modelling (OESA), Helmholtz-Centre for Environmental Research Ltd. - UFZ, D-04301 Leipzig, Germany; [k]School of Geography, Nanjing Normal University, 210023 Nanjing, China; [l]Key Laboratory of Virtual Geographic Environment, Ministry of Education of People's Republic of China, Nanjing Normal University, 210023 Nanjing, China; and [m]Deltares, 2600 Delft, The Netherlands

1. M. D. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018 (2016).
2. M. A. Janssen, C. Pritchard, A. Lee, On code sharing and model documentation of published individual and agent-based models. Environ. Model. Softw. 134, 104873 (2020).
3. V. Stodden, P. Guo, Z. Ma, Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. PLoS One 8, e67111 (2013).
4. M. Baker, Why scientists must share their research code. Nature (2016). https://doi.org/10.1038/nature.2016.20504.
5. N. P. Chue Hong, et al., FAIR principles for research software (FAIR4RS Principles). Research Data Alliance (2021). https://doi.org/10.15497/RDA00068.
6. C. Collberg, T. A. Proebsting, Repeatability in computer systems research. Commun. ACM 59, 62–69 (2016).
7. M. Baker, 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016).
8. Closing the climategate. Nature 468, 345 (2010).
9. K. M. Gustafsson, Learning from the experiences of the intergovernmental panel on climate change: Balancing science and policy to enable trustworthy knowledge. Sustainability 11, 6533 (2019).
10. J. Pipitone, S. Easterbrook, Assessing climate model software quality: A defect density analysis of three models. Geosci. Model Dev. 5, 1009–1022 (2012).
11. C. M. Barton et al., Call for transparency of COVID-19 models. Science 368, 482–483 (2020).
12. CESM, CESM metrics. https://www.cesm.ucar.edu/metrics. Accessed 1 August 2022.
13. CSDMS, CSDMS models by numbers. https://csdms.colorado.edu/wiki/CSDMS_models_by_numbers. Accessed 1 August 2022.
14. A. Morin et al., Shining light into black boxes. Science 336, 159–160 (2012).
15. V. Stodden, The scientific method in practice: Reproducibility in the computational sciences. SSRN Journal (2010). 10.2139/ssrn.1550193.
16. S. Stall et al., Make scientific data FAIR. Nature 570, 27–29 (2019).
17. N. Barnes, Publish your computer code: It is good enough. Nature 467, 753 (2010).
18. V. Grimm et al., The ODD protocol: A review and first update. Ecol. Modell. 221, 2760–2768 (2010).
19. E. Hutton, M. Piper, G. Tucker, The Basic Model Interface 2.0: A standard interface for coupling numerical models in the geosciences. J. Open Source Softw. 5, 2317 (2020).
20. F. Zhang et al., Interoperability engine design for model sharing and reuse among OpenMI, BMI and OpenGMS-IS model standards. Environ. Model. Softw. 144, 105164 (2021).