JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Editorial

Advancing Reproducibility and Open Data in Theoretical and Computational Chemistry

Cite This: https://doi.org/10.1021/acs.jctc.6c00733 Read Online

ACCESS | Metrics & More | Article Recommendations

We are pleased to present this joint editorial from the *Journal of Chemical Theory and Computation* (JCTC) and the *Journal of Chemical Information and Modeling* (JCIM), reflecting our shared commitment to advancing reproducibility, transparency, and open science in theoretical and computational chemistry.

The importance of sharing data, methods, and software is well established. In 2021, an editorial in *JCIM* (DOI: 10.1021/acs.jcim.0c01389) emphasized the central role of data and method availability in ensuring reproducibility in computational chemistry and biology. That need has only grown more urgent. The scale of computational data continues to expand, workflows are becoming more complex, and artificial intelligence (AI) and machine learning (ML) are increasingly integrated into chemical research. In this rapidly evolving landscape, clear and practical expectations for data and code availability are essential.

Open sharing benefits not only the broader scientific community but also the groups that generate these scientific assets. As articulated in the FAIR Guiding Principles, good stewardship ensures that data and related materials are findable, accessible, interoperable, and reusable. These practices make it easier to evaluate published results, reproduce key findings, and extend prior work to new systems and new questions. They also increase the long-term value of a study by preserving the underlying scientific record in forms that remain useful well beyond the original publication. When data, software, and workflows are shared in a clear and organized manner, they become durable contributions to the field.

To support these goals, the data and software needed to reproduce the key results of an original research manuscript should be made available at the time of manuscript submission whenever possible. This policy will take effect on May first 2026. Reviews, Perspectives, and Editorials are exempt. Original research articles should include a **Data and Software Availability Statement** that clearly indicates where the relevant data, code, and supporting materials can be accessed, ideally in machine-readable formats and through repositories that provide persistent identifiers.

An example statement is as follows:

The data underlying this study are available in the published article, the Supporting Information, and at [repository link]. The scripts and code used to generate and analyze the results are available at [repository link].

The materials needed to support reproducibility will vary across studies. Depending on the work reported, authors are encouraged to share, as appropriate, the following:

1. **Scripts and code** used to run calculations, generate inputs, process outputs, or analyze results. For new methods, the source code should be made accessible whenever possible, for example through GitHub, Zenodo, Figshare, or institutional repositories. Manuscripts should state clearly how the software can be accessed, whether openly or through a free or paid license.
2. **Representative input and output files** sufficient to reproduce key calculations or simulations. A clear repository structure and a brief README describing the workflow are strongly encouraged.
3. **System coordinates and related parameters**, including nonstandard basis sets, semiempirical parameters, force-field modifications, or other custom information required to define the systems under study.
4. **Key raw numerical data**, such as energies, free energies, observables, or other quantities underlying the principal conclusions.
5. **Representative restart files** when these are important for reproducing or continuing calculations.
6. **Simulation trajectories or representative structures**, particularly for molecular dynamics studies, together with the files and documentation needed to reproduce key simulations and interpret the resulting data.
7. **AI/ML models and training/validation/test information**, including model architecture, hyperparameters, relevant training, validation, and test data, and scripts or metadata necessary to reproduce the reported results. To compare different methods on a data set, the same training/validation/test splitting scheme should be enforced.
8. **Potential energy surfaces and related implementations**, including access to the data and relevant software needed to use newly reported surfaces in subsequent simulations.

Custom code central to the reported work should be made accessible whenever possible. Manuscripts should clearly specify whether the software used is commercial, open-source,

or in-house, together with the exact version employed; how in-house software can be accessed and what documentation is provided for external users; relevant dependencies, runtime environments, and, where appropriate, hardware specifications needed for reproducibility, as well as any licensing restrictions that apply.

We recognize that not every manuscript will require all these elements. Our aim is not to impose unnecessary burdens on authors, but to ensure that the information essential to reproducing the principal findings of a study is available in a form that is useful to reviewers, readers, and future researchers.

In some cases, legitimate constraints may limit full disclosure of data, models, or software, for example because of confidentiality, licensing restrictions, or third-party ownership. In such situations, the principal findings of the study should still be reproducible through sufficiently detailed methodological description and the sharing of all nonrestricted materials essential to the conclusions. Confidential or inaccessible components should not be indispensable to reproducing the main claims of the work.

We strongly encourage authors to use repositories that provide persistent identifiers such as DOIs or accession numbers. Repositories dedicated to research outputs, managed by sustainable institutions such as universities or professional societies, committed to long-term preservation and access, and transparent about terms of use are particularly valuable.

When code cannot be fully released, authors should still provide sufficient information for reviewers and readers to assess the reproducibility of the work. During peer review, this may include temporary access arrangements when feasible, preferably in a manner that preserves reviewer anonymity. For studies involving quantum computing hardware or specialized simulation platforms, authors should also report on the hardware and software environment in sufficient detail to support reproducibility on comparable systems.

NEW THEORETICAL DEVELOPMENTS

For manuscripts introducing new theoretical methods or frameworks, authors should provide complete and transparent descriptions of the underlying theory, algorithmic procedure, and implementation strategy. Key derivations should be included in the manuscript or the Supporting Information to the extent necessary for independent understanding, verification, and future adoption by the community.

LOOKING AHEAD

Reproducibility is a foundation of scientific progress. In theoretical and computational chemistry, where conclusions often depend on multistep workflows, specialized software, and increasingly large and heterogeneous data sets, reproducibility also depends on how scientific results are documented, preserved, and shared.

We recognize that meeting these expectations will require effort and that best practices will continue to evolve. Our intent is to foster a culture in which computational research is reported with the clarity and openness needed for others to verify, reuse, and build upon it. By working together as authors, reviewers, and editors, we can strengthen the rigor, accessibility, and long-term impact of the science published in JCTC and JCIM and help ensure that the discoveries reported today remain robust, reusable, and valuable for years to come.

Rommie E. Amaro orcid.org/0000-0002-9275-9553
Victor Batista orcid.org/0000-0002-3262-1237
Jochen Blumberger orcid.org/0000-0002-1546-6765
Yee Siew Choong orcid.org/0000-0001-5067-2073
Clémence Corminboeuf orcid.org/0000-0001-7993-2879
Zoe Cournia orcid.org/0000-0001-9287-364X
Qiang Cui orcid.org/0000-0001-6214-5211
Marco De Vivo orcid.org/0000-0003-4022-5661
Francesco A. Evangelista orcid.org/0000-0002-7917-6652
Yi Qin Gao orcid.org/0000-0002-4309-9376
Debashree Ghosh orcid.org/0000-0003-0726-7878
Xiao He orcid.org/0000-0002-4199-8175
Olexandr Isayev orcid.org/0000-0001-7581-8497
Syma Khalid orcid.org/0000-0002-3694-5044
Johannes Kirchmair orcid.org/0000-0003-2667-5877
John R. Kitchin orcid.org/0000-0003-2625-9232
Haiyan Liu orcid.org/0000-0002-5926-820X
Kevin J. Naidoo orcid.org/0000-0002-9898-3708
Duc Nguyen orcid.org/0000-0003-2215-0328
Ariane Nunes Alves orcid.org/0000-0002-5488-4732
Giulia Palermo orcid.org/0000-0003-1404-8737
Brett Savoie orcid.org/0000-0002-7039-4039
Thereza A. Soares orcid.org/0000-0002-5891-6906
Pratyush Tiwary orcid.org/0000-0002-2412-6922
Guowei Wei orcid.org/0000-0002-5781-2937
Xiao Zheng orcid.org/0000-0002-9804-1833
Tong Zhu orcid.org/0000-0001-7472-3736
Kenneth M. Merz, Jr. orcid.org/0000-0001-9139-5893
Laura Gagliardi orcid.org/0000-0001-5227-1396

Author Information

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.6c00733

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

ACKNOWLEDGMENTS

The authors thank the members of the Editorial Advisory Board and Early Career Board of JCTC and JCIM, and ACS staff who have offered useful comments on this editorial.

https://doi.org/10.1021/acs.jctc.6c00733
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX