

# Software Preservation, Stewardship, and Reuse

A Professional Guidance Statement of the American Meteorological Society

(Adopted by the AMS Council on 16 July 2021)

## Motivation:

Software is an essential component in driving scientific and technical advances in the atmospheric and oceanic sciences, leading to broader societal benefits. Society now relies upon software tools to assist in planning for daily life, improving the efficiency of economic activities, and saving lives when faced with pending natural disasters such as hurricanes. Modern numerical weather prediction (NWP) and ocean circulation models, which provide the foundation for environmental prediction, are essentially software products arising from decades of scientific research. As computational capacity and the complexity of observational networks increase, stewardship of software resulting from research is imperative in many cases. In order to build upon and further the knowledge that has been characterized within current software tools, the community is now expected to produce and curate software that is equitably accessible and easier to be reused by others. Equitable access to software that was used to discover the most recent research findings avoids wasteful duplication of efforts and provides an opportunity for any researcher to more easily build upon the work of others.

Accordingly, acceptance of the importance of open science principles [4] [13] [14] is growing throughout the scientific community. These principles intend to provide the foundation for more effective knowledge development and facilitate greater transparency in research processes [3]. Two critical components that support open science goals are open access to data and open access to research processes, including the relevant software components of those processes. In this context, “Software” refers to the executable set of instructions that direct a computer to create digital outputs [13] [15], where these outputs are used as evidence for the purposes of research or scholarship. There may be cases where simple numerical calculations or scripts can be easily explained in a published manuscript, eliminating the need to share related codes for future reuse.

The American Meteorological Society (AMS) is committed to supporting the principles of open science and has already developed guidance for data management [18] in support of its commitment of open access to data [6]. The purpose of this statement is to provide guidance to the AMS community on effective strategies to support software preservation, stewardship, reuse, and credit, including use cases where it is impractical to preserve and share large volumes of model output. Specifically, this statement will help AMS further community efforts to support open science and promote the principles, benefits, and effective practices being

developed by initiatives such as the Enabling FAIR Data Project [5] and FAIR for Research Software initiative [21] to AMS members and the broader atmospheric and ocean sciences community.

### Audience:

The following core principles and recommendations are intended to inform goals for standard practices by researchers in academia, government, and the not-for-profit and private sectors in the pursuit of open science. This is aligned with existing software preservation, stewardship, and reuse principles developed in the community, such as those described in the Earth Science Informatics Partners (ESIP) Software and Services Citations Guidelines and Examples [1], and the Future of Research Communications and e-Scholarship (FORCE11) Software Citation Principles [2].

### Specific Guidance:

The following principles and guidelines on software preservation, stewardship, and reuse serve as goals for the AMS community:

- ❑ **Open access to scientific research software.** Access to software developed and used by researchers is fundamental to advancing basic and applied science by building upon the work of others. Software should be “as open as possible, as closed as necessary” [9].
  - ❑ No matter how closed or protected the software is, the following recommendations apply:
    - ❑ Use a collaborative software development platform (e.g., GitHub or Bitbucket) to manage software code changes, and support public access capabilities when possible [7] [20].
    - ❑ Use a trusted preservation repository (e.g., Zenodo or Figshare) for long-term preservation and sharing of a software version snapshot used to support your research outcomes [12], or link to an archived snapshot of your software (e.g., by Software Heritage).
- ❑ **Well-documented software to optimize discovery, understanding, and support reuse and credit.** The creation and maintenance of robust software documentation and metadata supports discovery, reusability, credit, and knowledge growth. Archiving, preserving, stewarding, and accompanying software artifacts with community-accepted documentation and metadata ensures that the software will be fully discoverable and understandable for future research, serving the needs of science and society and giving

credit to the original developers [8] [11]. Following typical scholarly norms in citation, credit for software use and citation could then be included as an additional metric when evaluating academic or industry appointment, promotion, and recognition [19].

❑ **Curation, packaging, and access for current and legacy software versions.**

Equitable access to the version of software that was used to support a research finding is one method of facilitating research replication [4][10], avoiding wasteful duplication of effort and preventing bias in determining who can access that software. Consequently, current and major legacy versions of research software are best labeled, preserved, and made accessible through well-understood, standardized, and modern technical approaches [12]. Specifically:

- ❑ Assign a clear version number to your software in alignment with a specification such as semantic versioning [22].
- ❑ Ensure that your software is archivally preserved and provide a persistent identifier for the version you have used to support a research finding. Use persistent identifiers that are registered by international services (e.g., DataCite DOIs) to enable web-based discovery, access, and support for linking to the version of research software used in a publication. Repositories such as Zenodo and Figshare provide DOIs for archived software.
- ❑ Use common packaging systems, such as containers (e.g., Docker or Singularity), to enable straightforward installation on commonly used operating systems and computational platforms.

❑ **Software development, maintenance, and support costs.** Development and maintenance of research software that adheres to community best practice guidelines can be resource intensive. This may need to be considered when developing resource requirements for new projects and accounted for by sponsors in grant expectations.

❑ **Reasonable accommodation for intellectual property (IP) considerations and restrictive licenses.** IP concerns and restrictions exist in both public and private sector arenas.

- ❑ Assign a license that describes terms of software reuse and access [16][17].
  - ❑ Check with your institution and/or sponsor for guidance on choosing an appropriate software license.
  - ❑ If software cannot be publicly shared due to IP and/or licensing considerations, then, if possible, include a reference to a publication that describes the underlying logic and methods of software source code in the Terms of Use.

## References

- [1] Hausman, J., S. Stall, J. Gallagher, and M. Wu, 2019: Software and services citation guidelines and examples. ESIP, [https://esip.figshare.com/articles/journal\\_contribution/Software\\_and\\_Services\\_Citation\\_Guidelines\\_and\\_Examples/7640426/4](https://esip.figshare.com/articles/journal_contribution/Software_and_Services_Citation_Guidelines_and_Examples/7640426/4).
- [2] Smith, A. M., D. S. Katz, K. E. Niemeyer, and FORCE11 Software Citation Working Group, 2016: Software citation principles. *PeerJ Comput. Sci.*, **2**, e86, <https://doi.org/10.7717/peerj-cs.86>.
- [3] National Academies of Sciences, Engineering, and Medicine, 2019: *Reproducibility and Replicability in Science*. National Academies Press, 256 pp., <https://doi.org/10.17226/25303>.
- [4] National Academies of Sciences, Engineering, and Medicine, 2018: *Open Science by Design: Realizing a Vision for 21st Century Research*. National Academies Press, 232 pp., <https://doi.org/10.17226/25116>.
- [5] Stall, S., and Coauthors, 2018: Advancing FAIR data in Earth, space, and environmental science. *Eos*, **99**, <https://doi.org/10.1029/2018EO109301>.
- [6] American Meteorological Society, 2019: Full, open, and timely access to data. <https://www.ametsoc.org/index.cfm/ams/about-ams/ams-statements/statements-of-the-ams-in-force/full-open-and-timely-access-to-data/>.
- [7] Stodden, V., J. Seiler, and Z. Ma, 2018: An empirical analysis of journal policy effectiveness for computational reproducibility. *Proc. Natl. Acad. Sci.*, **115**, 2584–2589, <https://doi.org/10.1073/pnas.1708290115>.
- [8] Chue Hong, N. P., and Coauthors, 2019: Software citation checklist for authors (version 0.9.0). Zenodo, <https://doi.org/10.5281/zenodo.3479199>.
- [9] European Commission, 2020: H2020 online manual for data management. [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm).
- [10] EarthCube RCN, 2020: "What about model data?" Determining best practices for archiving and reproducibility. <https://modeldatarcn.github.io/>.
- [11] Katz, D. S., and Coauthors, 2021: Recognizing the value of software: A software citation guide [version 2; peer review: 2 approved]. *F1000Research*, **9**, 1257, <https://doi.org/10.12688/f1000research.26932.2>.
- [12] Chue Hong, N. P., and Coauthors, 2019: Software citation checklist for developers (version 0.9.0). Zenodo, <https://doi.org/10.5281/zenodo.3482769>.

[13] Lamprecht, A.-L., and Coauthors, 2020: Towards FAIR principles for research software. *Data Sci.*, **3**, 37–59, <https://doi.org/10.3233/DS-190026>.

[14] Wilkinson, M. D., and Coauthors, 2016: The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*, **3**, 160018, <https://doi.org/10.1038/sdata.2016.18>.

[15] Katz, D. S., and Coauthors, 2016: Software vs. data in the context of citation. *PeerJ Preprints*, **4**, e2630v1, <https://doi.org/10.7287/peerj.preprints.2630v1>.

[16] Choose an open source license, <https://choosealicense.com/>.

[17] Five recommendations for FAIR software, <https://fair-software.nl/>.

[18] American Meteorological Society, 2019: Best practices for data management. <https://www.ametsoc.org/index.cfm/ams/about-ams/ams-statements/statements-of-the-ams-in-force/best-practices-for-data-management/>.

[19] Moher, D., F. Naudet, I. A. Cristea, F. Miedema, J. P. A. Ioannidis, and S. N. Goodman, 2018: Assessing scientists for hiring, promotion, and tenure. *PLoS Biol.*, **16**, e2004089, <https://doi.org/10.1371/journal.pbio.2004089>.

[20] Krafczyk, M. S., A. Shi, A. Bhaskar, D. Marinov, and V Stodden, 2021: Learning from reproducing computational results: Introducing three principles and the *Reproduction Package*. *Philos. Trans. Roy. Soc.*, **379A**, 20200069, <https://doi.org/10.1098/rsta.2020.0069>.

[21] Research Data Alliance, 2021: FAIR for Research Software (FAIR4RS) WG. <https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg>.

[22] Preston-Werner, T., 2013: Semantic versioning 2.0.0. <https://semver.org/>.

[This statement is considered in force until July 2026 unless superseded by a new statement issued by the AMS Council before this date.]