Credit where credit is due: Data and software in the space weather community

S. K. Morley[1], H. Liu[2], B. A. Carter[3], J. L. Gannon[4], and N. Lugaz[5]

[1]Space Science and Applications, Los Alamos National Laboratory, Los Alamos, USA
[2]Department of Earth and Planetary Science, Kyushu University, Fukuoka, Japan
[3]STEM College, RMIT University, Melbourne, Australia
[4]Computational Physics, Inc., Boulder, CO, USA
[5]Department of Physics and Astronomy, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH, USA

Key Points:

- Open and accessible resources now enable FAIR science to an unprecedented degree
- Open data and software enable research to be built upon while providing credit to originators of nontraditional research output
- Restrictions can remain in applied work and the editors aim to help navigate the balance

Corresponding author: Steven K. Morley, `smorley@lanl.gov`

Abstract

This editorial aims to improve awareness of the current best practices in open research, and stimulate discussion on the practical implementation of AGU's data and software policy in key areas of space weather research. We also further aim to encourage authors to take additional steps to ensure clear credit to all contributors to the work, whether that is underlying data, key software, or direct contributions to the manuscript.

Over recent decades, AGU has established and developed data and software polices for authors that strive to make published research open and reproducible (Hanson & van der Hilst, 2014). These policies (`https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors`) aim to ensure that the data and key software required to evaluate and build on the published work are available for readers during both peer review and after publication. This also highlights the need to recognize and credit the providers and maintainers of data and software.

The space weather community occupies the application-oriented edge of space research, and as such engages both directly and indirectly with forecast centers, industry, government and other end users. Work at this interface heightens the importance of robust and reproducible science based transparent approaches. Applying AGU's data and software policies can be challenging for applied research, especially in cases using data from systems that have proprietary, commercial, or national security concerns. For example, for satellite anomalies the anomaly details may be considered commercially sensitive, while technical specifications may be additionally controlled by local export control laws. Similarly for power grid impacts, while some geomagnetically induced current (GIC) data are publicly available, these data and infrastructure details required for detailed simulation and interpretation of impacts on power flow and systems are often controlled. Publication of results that use restricted data or software is still of significant value and is supported by *Space Weather* in cases where the data and software policies might otherwise hinder the path to publication.

In a previous editorial, Hapgood and Knipp (2016) wrote about open research, data availability, and data citation in the context of space weather research. As the environment around this crucial topic continues to evolve, we provide an update and speak to some additional considerations for open research. A number of developments over recent years have changed that environment. In particular, access to relatively large-scale data and software services has become widespread, free of charge, and relatively user-friendly. Large-scale data archival and discoverability services are now available without cost to the user, and the same is true for version control of open-source software.

Recently the momentum of open and reproducible science has coalesced around the FAIR (Findability, Accessibility, Interoperability, and Reuse) guidelines (Wilkinson et al., 2016). These guidelines provide key principles for scientists to follow when performing and reporting on their science. This editorial aims to improve awareness of the current best practices and stimulate discussion on the practical implementation of the policy in key areas of space weather research. We also further intend to encourage authors to take additional steps to ensure clear credit to all contributors to the work, whether that is underlying data, key software, or direct contributions to the manuscript.

In addition to the AGU guidance, numerous papers exist that aim to help scientists put these principles into practice in their work (e.g., Alston & Rick, 2021). We note that, especially in the context of applied work including commercial or government stakeholders, the FAIR guidelines may be challenging to fully implement for any individual piece of work. However, there is movement across journals including Space Weather and funding agencies (e.g., NASA's Transform to Open Science (TOPS) initiative; `https://science.nasa.gov/open-science/transform-to-open-science`). There is also widespread, but not universal, support for both open data and open software (National Academies of

Sciences & Medicine, 2018, especially Appendix C). However, while some of these concerns remain to be navigated in implementing the ideals of open science across publishers and funding agencies, Space Weather requires that *where possible* these principles are adhered to and, as noted earlier, the editors strive to work with authors on a case-by-case basis to balance ideals and practicality.

Data sets, especially large-scale (e.g., long-term satellite missions) often do not have DOIs for their data products, and developing this infrastructure requires significant effort. This includes scientists with significant knowledge of the data and the relevant metadata standards and persistent identifier generation. One example often that has wide adoption is the SPASE metadata model (Roberts et al., 2018) with associated data access through a flexible interface such as the Heliophysics Application Programmer's Interface (HAPI; Weigel et al., 2021). In cases where data providers and archival services do not yet provide digital object identifiers (Chandrakar, 2006) or similar persistent identifiers (Lubas et al., 2022), other information can typically be leveraged by authors to ensure the highest chance of reproducibility. For example, specific file names and versions can be provided for each data product used. URLs should be provided for individual data products where possible instead of landing pages for a mission. Where data have been generated for a particular project, these should be submitted to a service that will both host the data and assign a DOI that can be cited in text. These considerations also apply to software, where key software should be cited if possible. Many community software libraries and tools are open-source and have both open development (e.g., on a platform like GitHub) and citeable releases via an archival service like Zenodo. Of course, many data sets and software packages have peer-reviewed articles describing them – in some cases, especially for legacy data and software, this is the primary description – and these should be cited in addition to the software itself.

Interestingly, the concept of persistent identifiers has been extended to individual researchers who now can be uniquely identified using an identifier like the Open Researcher and Contributor ID (ORCID; Butler, 2012), which is supported by AGU journals. In addition to FAIR and ORCID, the open science ecosystem includes the Contributor Roles Taxonomy (CRediT; see Brand et al., 2015), which provides a vocabulary for clearly identifying contributor roles. AGU journals also support CRediT for explicitly stating author contributions to a manuscript. [1]

Finally, whether data or software are under consideration, licensing must be considered to ensure that research products can be used by their intended audience. For example, a numerical model released under a "copyleft" style license cannot subsequently be used within a predictive system that uses a permissive license. Licenses restricting who may use the data or code are typically not considered open, and commercial use restrictions can have both benefits and drawbacks (Fang et al., 2022). Large mission data sets have traditionally included "rules of the road" [2] that function similarly to a license, though are typically not crafted with the exact same aims in mind. However, explicit licensing is recommended to ensure that the data have clear terms of use and intellectual property (IP) protection. Not all licences are compatible with each other, and license compatibility can also represent a hurdle to building on work that otherwise meets the ideals of open science. More permissive licenses are most likely to allow interoperability and compatibility between different data sets and software systems and are recommended for meeting open science ideals. For data, archived presentations, etc. the Creative Commons CC-BY and CC0 are examples of permissive licences. For software, permissive licenses approved by the Open Source Initiative are good examples. Licensing of data, software, or other scientific outputs may require coordination with the en-

[1] https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Text-requirements

[2] e.g., https://www.sws.bom.gov.au/World_Data_Centre/1/5/3; https://lasp.colorado.edu/galaxy/display/MFDPG/1.2+MMS+and+FPI+Rules+of+the+Road

tities employing the contributors and/or funding the research (Appendix B of National Academies of Sciences & Medicine, 2018), as employment and funding agreements typically specify the owner of IP rights for any given work.

We encourage our community to work towards research that is accessible to all and gives credit to all involved in the process, whether that is data collection, software development, or the scientific work directly leading to submitted manuscripts.

1 Open Research

No data or analysis software were generated or used in the preparation of this manuscript.

Acknowledgments

Portions of this work by SKM were performed under the auspices of the US Department of Energy and reflect personal opinions.

References

Alston, J. M., & Rick, J. A. (2021). A beginner's guide to conducting reproducible research. *The Bulletin of the Ecological Society of America*, *102*(2), e01801. Retrieved from `https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/bes2.1801` doi: https://doi.org/10.1002/bes2.1801

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*, *28*(2), 151-155. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1087/20150211` doi: https://doi.org/10.1087/20150211

Butler, D. (2012, May 01). Scientists: your number is up. *Nature*, *485*(7400), 564-564. Retrieved from `https://doi.org/10.1038/485564a` doi: 10.1038/485564a

Chandrakar, R. (2006). Digital object identifier system: an overview. *The Electronic Library*.

Fang, T.-W., Kubaryk, A., Goldstein, D., Li, Z., Fuller-Rowell, T., Millward, G., ... Babcock, E. (2022). Space weather environment during the spacex starlink satellite loss in february 2022. *Space Weather*, *20*(11), e2022SW003193. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022SW003193` (e2022SW003193 2022SW003193) doi: https://doi.org/10.1029/2022SW003193

Hanson, B., & van der Hilst, R. (2014). Agu's data policy: History and context. *Eos, Transactions American Geophysical Union*, *95*(37), 337-337. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014EO370008` doi: https://doi.org/10.1002/2014EO370008

Hapgood, M., & Knipp, D. J. (2016). Data citation and availability: Striking a balance between the ideal and the practical. *Space Weather*, *14*(11), 919-920. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016SW001553` doi: 10.1002/2016SW001553

Lubas, R. L., Koskas, M., Committee, I. B. S. S., Riva, P., Guerrini, M., Häusner, E.-M., ... et al. (2022, Jun). *Common practices for national bibliographies in the digital age.* International Federation of Library Associations and Institutions (IFLA). Retrieved from `https://repository.ifla.org/handle/123456789/2001`

National Academies of Sciences, E., & Medicine. (2018). *Open source software policy options for nasa earth and space sciences.* Washington, DC: The National Academies Press. Retrieved from `https://nap.nationalacademies.org/catalog/25217/open-source-software-policy-options-for-nasa-earth`

-and-space-sciences doi: 10.17226/25217

Roberts, D. A., Thieman, J., Génot, V., King, T., Gangloff, M., Perry, C., … Hess, S. (2018). The spase data model: A metadata standard for registering, finding, accessing, and using heliophysics data obtained from observations and modeling. *Space Weather*, *16*(12), 1899-1911. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018SW002038 doi: 10.1029/2018SW002038

Weigel, R. S., Vandegriff, J., Faden, J., King, T., Roberts, D. A., Harris, B., … Martinez, B. (2021). Hapi: An api standard for accessing heliophysics time series data. *Journal of Geophysical Research: Space Physics*, *126*(12), e2021JA029534. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JA029534 (e2021JA029534 2021JA029534) doi: https://doi.org/10.1029/2021JA029534

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016, Mar 15). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. Retrieved from https://doi.org/10.1038/sdata.2016.18 doi: 10.1038/sdata.2016.18