



## OPEN ACCESS

## EDITED BY

Katariina Nykyri,  
Embry–Riddle Aeronautical University,  
United States

## REVIEWED BY

Luke Barnard,  
University of Reading, United Kingdom  
Andrew Dimmock,  
Institute for Space Physics (Uppsala),  
Sweden

## \*CORRESPONDENCE

Alexa J. Halford,  
Alexa.J.Halford@NASA.gov

## SPECIALTY SECTION

This article was submitted to Space  
Physics,  
a section of the journal  
Frontiers in Physics

RECEIVED 04 October 2022

ACCEPTED 08 November 2022

PUBLISHED 05 December 2022

## CITATION

Halford AJ, Chen TY and Rastaetter L  
(2022), Data needs to be a priority.  
*Front. Phys.* 10:1061681.  
doi: 10.3389/fphy.2022.1061681

## COPYRIGHT

© 2022 Halford, Chen and Rastaetter.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Data needs to be a priority

Alexa J. Halford<sup>1\*</sup>, Thomas Y. Chen<sup>2</sup> and Lutz Rastaetter<sup>1</sup>

<sup>1</sup>Heliophysics Division, Goddard Space Flight Center, NASA, Greenbelt, MD, United States, <sup>2</sup>The Fu Foundation School of Engineering and Applied Science, Columbia University, New York, NY, United States

Findability, Accessibility, Interoperability, and Reusability (FAIR) data are essential to heliophysics and all scientific research. The principles of FAIR data ensure the reusability and findability of data, as well as its long-term care. The goal is that data are accessible for the ongoing discovery and verification process and can be used on their own or with newly generated data in future studies leading to innovations. With the onset in the previous decades of NASA and other agencies requiring mission data to be open to the public, heliophysics has already made great strides toward FAIR data and benefited from these efforts. Continued improvements in our metadata, data archives, and data portals and the addition of DOIs for data citation will ensure data will be FAIR, enabling further scientific discoveries, reproducibility of results, longitudinal studies, and verification and validation of models. Currently, not all the data collected are findable and on open networks or archives, and not all data on archives have DOIs. Within this study, we make recommendations to prioritize resources needed to satisfy FAIR data principles, treating them as a fundamental research infrastructure rather than a simple research product.

- Data collection, preparation, archiving, and accessibility need to be a priority.
- Data collection, preparation, archiving, and accessibility need dedicated and sustained funding support.
- Data need to be accessible through investment in infrastructure: tools to access and read the data and personnel to maintain these data and IT infrastructure.
- Data need to be collected in sustained ways to enable further science and, specifically, model validation efforts.

## KEYWORDS

open science, data management preservation and rescue, data and information governance, metadata, portals and user interfaces, validation, reproducibility, web services

## 1 Introduction

Data are the foundation of good science. This is true whether data are observations of the physical world or output from computer simulations. Data must be collected carefully, prepared consistently, interpreted in an appropriate context, archived, and made accessible for reproducibility and subsequent research. Moreover, in recent years, data

have been made interoperable with other datasets as the most cutting-edge science is interdisciplinary and spans multiple physical domains.

The growing awareness of Findability, Accessibility, Interoperability, and Reusability (FAIR) data principles (<https://www.go-fair.org/fair-principles/>) is an acknowledgment of these truths within the broader scientific community [1]. However, consistent application of these principles is a scientific challenge of its own, and different disciplines have made more and less progress toward these important goals [2–4]. Several FAIR data principles work around ensuring data and metadata are properly developed using standard vocabularies, adequately archived, and open, which ensure that data can be easily found by a human or machine. Heliophysics is on its way toward ensuring FAIR data by adopting metadata standards such as the Space Physics Archive Search and Extract (SPASE) (<https://spase-group.org/data/>), which is a Committee on Space Research (COSPAR) recommendation.

The following is a list of recommendations we believe will, if followed, bring heliophysics up to par with other fields that have been working longer on implementing FAIR data principles (we should not assume they have solved all challenges [5]). Each recommendation includes several suggested actions intended to help realize the stated goals.

This study was originally written in response to the 2024 heliophysics decadal survey run by the National Academies in the United States. Previous decadal surveys and other national academy reports have been used by many space agencies within the United States to help form their strategic plans for the coming decade [6–8]. Thus, this study has a more US-centric view with recommendations aimed at what institutions in the United States might be able to adopt to help ensure that data are a priority. However, we feel that many of these recommendations apply to other institutions, which can be taken on an individual level.

## 2 Recommendation 1: Data needs to be a priority

Historically, the expense and time of data production have led to a relatively low priority of what is now known as FAIR data principles [2]. Data production of higher-level and inter-calibrated products, ensuring proper metadata is developed, and archiving often occur at the end of missions when there is little or no funding. However, it is more cost-effective to maintain datasets than launch new satellite missions, develop new supercomputer models, or re-run existing models. Additionally, there are long-period oscillations in solar activity (11-year sunspot cycle [9–15]) and even slow secular variations in Earth's main magnetic field where we have yet to obtain or prioritize longitudinal sets of these and

similar data. This lack of data creates roadblocks for researchers interested in space climatology. Finally, while sensors and models will usually improve with better engineering and more sophisticated algorithms, good data stewardship is essential to track the evolution of data and model quality over time, thus assessing the quality of the science.

Heliophysics-related data stewardship must be a priority in its own right, on par with, but independent from, data collection and interpretation.

Many scientists and programs are dedicated to making scientific data FAIR [1–3]. However, these efforts have been hampered by outdated academic expectations and practices such as “publish or perish,” not to mention insufficient and unreliable funding. We offer possible actions that will mitigate some of the current challenges.

### 2.1 Action 1 toward FAIR data being a priority

We acknowledge that the FAIR data landscape changes quickly and rapidly over a decade. Thus, we encourage this topic to be revisited regularly, more than once every 10 years.

### 2.2 Action 2 toward FAIR data being a priority

Make Space Physics Data Facility (SPDF) a Goddard Heliophysics Tier 1 capability to match the CCMC priority and thus make it easier to ensure funding and connection between model and observational data validation. The SPDF and similar archives/portals housed at NASA Goddard have become nodes where the community can freely find many space physics datasets, especially satellite, rocket, and balloon-based data. SPDF currently has the personnel to help ensure all data follow a basic set of FAIR standards, such as file type and metadata [1–4]. It has been so successful that other groups also have data access through them, including NOAA and ground-based indices. We suggest continued and further funding of archive centers such as SPDF to expand and provide portal access to other data centers and archives, such as SuperMag and Madrigal [16,17].

### 2.3 Action 3 toward FAIR data being a priority

Ensure similar funding/resources for other agencies with a portability/portal to connect all agency and industry/academic data. While SPDF covers much of NASA heliophysics data, we see a symmetric need at other agencies and institutions that

develop and curate data portals [18,19]. Institutions need sufficient and reliable funding to maintain their own data archives and accessible portals to access the data. Without reliable and consistent funding, data are often lost and sometimes not recoverable. Multiple repositories exist in redundancy in case of loss.

### 3 Recommendation 2: Data needs dedicated funding

The dedicated new funding is necessary to ensure the lasting infrastructure for data archival and access. Data archival can be achieved through multiple platforms and, ideally, will use multiple platforms to help ensure redundancy. All options take expertise, equipment, and upkeep.

As new dedicated funding may not be immediately available to ensure the lasting infrastructure necessary for data archival and access, value judgments will need to be made. Currently, there are many data archival locations. Often data from missions in short and, sometimes, long terms are housed at individual institutions. We have seen in the past how this can lead to data loss when people retire, leave the field, or need that storage space for something else. Therefore, a new mission is needed to gain data after its loss, which often will cost more than just maintaining the data. As a field, we have lost too much already. Data archival plans can be included in the proposal, such as the data management plan for ROSES calls. Thus, data processing, publication, and archiving all need to be incorporated into the proposed budget. This may lead to less funding for science research in the immediate term but will enable more scientific studies and gains in the long run.

#### 3.1 Action 1 toward FAIR data being funded

Similar funding/resources for agencies with a portability/portal to connect all agency and industry/academic data. As we raise data stewardship as a priority across all agencies, this effort needs to be supported through funding the infrastructure and maintenance, as well as community access capabilities, for example, development of APIs or standard code libraries for data access, load, and piloting routines [4,20,21].

#### 3.2 Action 2 toward FAIR data being funded

Ideally, there will be multiple data repositories for redundancy. Often small and older missions have data housed on antiquated platforms (e.g., CDs or floppy disks) or on an

institutional or personal system such as Dropbox or a university server. These data are too often lost when people change jobs, retire, or systems fail. We recommend that at least one repository be government-based as the government, more than academia or industry, has an obligation to ensure data archive purposes and data accessibility to the public, ensuring open science is achieved. The multi-platform approach would utilize government-owned and commercial platforms, including cloud technologies. While we expect much innovation from commercial entities, we recommend that the government also considers providing services similar to basic cloud-based platforms and other collaborative tools, which would ensure further accessibility to data and computing to the public (e.g., [git.gov](https://git.gov)), specifically for those who can not afford the costs of commercial tools, cloud computing, and storage.

#### 3.3 Action 3 toward FAIR data being funded

We see a growing need for funding small projects to port data to the archives. We acknowledge that this, in part, already exists within the NASA proposal system but may be lacking at other space funding institutions. However, the NASA data calls are still not well-known throughout the community. We see data collection, processing, archiving, and analysis, all working toward new physical understanding, as an entire ecosystem that needs to be supported throughout its lifecycle by funding agencies. As an example of how the different calls could work, we will use an example from the existing NASA proposal call structure.

One such expansion of the current NASA ROSES Heliophysics Data Environment Enhancements (HDEE) calls could be a dedicated call for successfully flown low-cost access to space (LCAS) missions. Once the mission flies, the team could propose making the data more easily accessible and available to the public. Although many LCAS missions work to make their data public, they are run on shoestring budgets, meaning that items and goals at the end of the missions are less likely to be funded and achieved in full. Therefore, while the data may become technically public, it is in a format that is difficult to be used by non-team or even non-instrument providers. A NASA ROSES Guest Investigator (GI) type program, where data stewards, data specialists, or data historians apply and work with the instrument team to further process the data, includes proper metadata, and basic read, load, and plot routines would greatly enable accessibility and further science returns from these low-cost missions [4,22,23]. Larger missions may also benefit from similar expansions of their calls to include specific proposals for data specialists to work with mission teams to help ensure data accessibility and usability, ensuring adherence to best practices and current standards.

### 3.4 Action 4 toward FAIR data being funded

Although the above has focused on observational data, the same recommendations hold true for model data. Running a large simulation is costly financially and computationally and impacts our carbon footprint. Retaining model runs and improving data will enable better and faster model data comparisons and validation activities, as well as including more model data and global views for data-driven studies.

## 4 Recommendation 3: Data needs to be accessible

Software libraries, interactive applications, and web-based application program interfaces (web-APIs) are different tools necessary for accessing data. As popular tools change and update, expertise and resources are needed to ensure continued access to data. Many groups have identified an additional need for repositories or portals, which may not house data but point users or grab data for users through their interface [22,23]. As popular tools change and are updated, expertise and resources are needed to ensure continued access to the data. We can look to other fields for successful examples, such as IRIS in seismology ([https://www.iris.edu/hq/about\\_iris](https://www.iris.edu/hq/about_iris)) [24,25].

### 4.1 Action 1 toward data being accessible

There is a strong need to better connect the different data portals/observatories/viewers/archives (e.g., SPDF, Helioviewer, madrigal, Supermag, TREX, and AuroraX) [16,17,23,26–28]. Many of the current data archives and portals were built around one sub-field of heliophysics, such as solar, ionosphere, or magnetosphere, and sometimes even smaller sub-fields, such as radiation at aviation altitudes. As we look toward a future with more transdisciplinary science, we see a growing need to have these data portals more interconnected. However, this recommendation can only come once the data archives and repositories and their respective portals are funded and sufficiently operational with plans on how to interconnect these facilities, as suggested in Sections 3.1 and 3.2.

### 4.2 Action 2 toward data being accessible

Standardized data formats enable easier access and increased use of data [29]. ASCII, comma-separated variables, are human-readable data types that will always be useful but often take up much space. Compressed file types or binaries are vital for our ever-growing data. We suggest moving toward Hdf5 or NetCDFs as they have a larger

user database and, thus, a wider developer base for reading routines [30–32]. This wider user base also increases the needed baseline support level and timeliness of updates in the future.

### 4.3 Action 3 toward data being accessible

With historical datasets and the likelihood of new data formats in the future, there is a clear need for continued development and maintenance of converters between data types. Maintaining converters between data types ensures the continued use and science gains from the data in the future. We also encourage data archives to update data into new currently standard data types.

### 4.4 Action 4 toward data being accessible

Maintaining data readers for multiple code languages is necessary to ensure their continued use within the research community. The research community uses multiple different types of coding languages, and ensuring easier access to data will enable further science returns. Providing readers also reduces potential errors in correctly reading data and thus correctly processing and interpreting it.

Metadata is necessary to ensure data findability [4]. However, producing and developing metadata is time-consuming, requiring skills that not all researchers or data providers continually cultivate. As our field grows and expands, it is necessary to ensure metadata standards are maintained and improved so that our data can be easily accessed and used. This may also necessitate encouraging the development of new professional titles and the cultivation of new skill sets in our field including data curators and data specialists.

### 4.5 Action 5 toward data being accessible

Simplify metadata with clear ways to crowd-source and improve it with continued use. Simplifying the initial basic metadata needed reduces the barrier of entry for making data public. Providing a way for others to contribute to the further updates of metadata will allow the metadata to be updated, as well as providing insights into what science phenomena the data may be used to study, which may not have been considered by the data providers. For example, the initial metadata may be a paired-down version of what is expected from SPASE, such as instrument, PI, and units. Then, crowd sourcing may add expected min/max values, phenomenon types, wave types, or other useful metadata that a user can contribute.

### 4.6 Action 6 toward data being accessible

Develop wrappers or code packages to help produce basic metadata and identify what metadata and common format,

among others, are missing. This would help data providers create and know what information is needed and ease the data transition from the data provider to data archives/repositories such as SPDF. For example, SPASE exists as the protocol for metadata, but it is not easy to use. There is a large learning curve for new people to make their datasets SPASE-compliant. It would be fantastic to have a tool to help someone conform to the SPASE protocol.

## 4.7 Action 7 toward data being accessible

Maintaining, archiving, and ensuring accessibility to data take time and effort, which must be recognized. We encourage developing new metrics or using current metrics to give credit to these activities [33–35]. As hiring, promotion, and awards often rely in part on quick metrics such as the H-index, there is a growing need to find ways to give credit to this unseen and often unacknowledged labor and skills. We encourage institutions to help promote and aid in the following:

- Further the culture to accurately cite data (e.g., SPDF providing DOIs on the screen where data are chosen to be plotted/downloaded).
- Encourage more studies on data descriptions. We suggest a journal or journal article type like those in JOSS (<https://joss.theoj.org/>), approximately one pager describing the data and metadata to advertise the new data availability. This type of peer review journal article for the data can check for such items as follows: do the data exist, do they have appropriate metadata, and are they accessible?

## 5 Recommendation 4: Data needs to be collected with global model validation and systems science as priorities

Although we have been visiting space for over 70 years and remotely observing space from Earth and the lower atmosphere for a few hundred, we have not adequately sampled all regions of space that our models and theory cover. We need more data and supporting missions that prioritize systems science and model validation to advance our understanding of the dynamics and coupling across the heliosphere. As access to space is cost-prohibitive, we should ensure that we make the most of historical datasets [36–40].

### 5.1 Action 1 toward more data

Benchmark activities need to become more formalized and standardized within the research community. Time intervals for different phenomena intervals, such as quiet, solar storm,

geomagnetic storm, ionospheric impacts, and atmospheric impacts, need to be identified and agreed upon by the community to provide a standard set of data for model validation activities [36,41–45]. Groups such as the modeling, methods, and validation group through Geospace Environment Modeling (GEM) ([https://gem.epss.ucla.edu/mediawiki/index.php/RG:\\_Modeling\\_Methods\\_and\\_Validation](https://gem.epss.ucla.edu/mediawiki/index.php/RG:_Modeling_Methods_and_Validation)) and the International Space Weather Action Teams (ISWAT) COSPAR (<https://www.iswat-cospar.org/iswat-cospar>) are two groups within the community who are working toward developing these benchmarks.

### 5.2 Action 2 toward more data

Expand or create new mission lines to focus on questions around increasing data for model validation. Current mission lines prioritize fundamental science questions. While needed validation for models can be shoehorned into this framework, we believe better validation efforts will be achieved if this is prioritized. We envision this as a new funding line with a budget of at least that of explorers and potentially much larger as multiple well-instrumented platforms may be necessary to gather sufficient data for the validation efforts.

### 5.3 Action 3 toward more data

Model data need to be archived in a similar way to observable datasets. Running models can be both time- and funding-intensive. Thus, much work through metadata groups, such as SPASE, has worked to be inclusive of the specific needs of model data for their metadata. Additionally, we need to ensure that models can be compared with observational data for validation and to improve studies that make use of both observable and model data, given our scarcity of observational sites. Thus, we need to increase incentives to have model outputs for comparison with observational assets in similar physical units to observational data.

- Parameters such as Phase Space Density  $L^*$  are intrinsic to models but not to our observational capabilities. Thus, to better calculate errors within our measurements, observations, and analysis, we encourage transforming the model outputs into observational parameters instead of converting observational data into model outputs.
- Likewise, we need to treat model data similarly to observational data by working toward collecting and analyzing the model data in a similar fashion to the observed data. This may include the following:
  - Integration time to collect particle observations and field-aligned current (FAC) maps, among others.
  - Integration of climatology *versus* single model run instance (chorus wave occurrence and particle loss)

## 5.4 Action 4 toward more data

A new research line is needed with similar funding levels to the Guest Investigator Program, focusing on using historical data to address ongoing science questions. Historical data are incredibly valuable to address the primary science questions of the original mission and new science questions long after the end of the mission ended. A new funding line focused on using historical data will further extend science from these past missions and extend the usefulness of the data archives and portals.

## 6 Conclusion

As we move into the coming decade, data need to be a priority, FAIR, and with dedicated funding. They need to be collected with validation and system science as a priority. Easily accessible FAIR data have a positive impact on the amount and quality of science that can be completed by a scientific community [46–50]. It is time for Space Physics and Space Weather, where we have fundamental dynamics in the system that occur over long timescales, necessitating long data storage, to prioritize data and data archiving. These four top-level recommendations and the suggested actions will help ensure data are available to answer the science questions, develop space weather tools, and help with our validation needs for the coming decade.

Our final set of high-level recommendations for the next 10 years are as follows:

- Data need to be a priority.
- Data need dedicated funding.
- Data need to be accessible.
- Data need to be collected with global model validation and systems science as priorities.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

## References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* (2016) 3. doi:10.1038/sdata.2016.18
2. Stall S, Yarmey L, Cutcher-Gershenfeld J, Hanson B, Lehnert K, Nosek B, et al. Make scientific data FAIR. *Nature* (2019) 570:27–9. doi:10.1038/d41586-019-01720-7
3. Kinkade D, Shepherd A. Geoscience data publication: Practices and perspectives on enabling the fair guiding principles. *Geosci Data J* (2022) 9: 177–86. doi:10.1002/gdj3.120
4. Roberts DA, Thieman J, Génot V, King T, Gangloff M, Perry C, et al. The space data model: A metadata standard for registering, finding, accessing, and using heliophysics data obtained from observations and modeling. *Space Weather* (2018) 16:1899–911. doi:10.1029/2018SW002038
5. Kelbert A. Science and cyberinfrastructure: The chicken and egg problem. *Eos Trans AGU* (2014) 95:458–9. doi:10.1002/2014eo490006
6. Council NR. *Solar and space physics: A science for a technological society*. Washington, DC: The National Academies Press (2013). doi:10.17226/13060

## Author contributions

AH wrote the initial draft of the paper. All co-authors helped with editing the paper.

## Funding

AH and LR were supported by the Space Precipitation Impacts project at Goddard Space Flight Center through the Heliophysics Internal Science Funding Model.

## Acknowledgments

Valuable input and feedback were offered by E. Joshua Rigler, USGS. The authors would like to acknowledge the MMV Resource Group from GEM and the GEM Community for discussions about this important topic at the 2022 summer GEM workshop. The authors would like to thank overleaf, which aided our ability to collaborate. The positions, experiences, and viewpoints expressed in this work are those of the authors as scientists in the space research community and are not the official positions of their employing institutions.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

7. National Academies of Sciences, E. and Medicine. *Report series: Committee on solar and space physics: Heliophysics science centers*. Washington, DC: The National Academies Press (2017). doi:10.17226/24803
8. NASA. *Heliophysics living with a star program, 10-year vision beyond 2015* (2015). Available from: <https://tinyurl.com/LWS-10-year-plan> (Accessed October 27, 2022).
9. Kirkwood D. On the periodicity of the solar spots. *Proc Am Philos Soc* (1869) 11:94–102.
10. Bloxham J, Gubbins D. The secular variation of Earth's magnetic field. *Nature* (1985) 317:777–81. doi:10.1038/317777a0
11. Clette F, Svalgaard L, Vaquero JM, Cliver EW. Revisiting the sunspot number. In: *Space sciences series of ISSI*. New York: Springer (2015). p. 35–103. doi:10.1007/978-1-4939-2584-1\_3
12. Alken P, Thébaud E, Beggan CD, Nosé M. Special issue “international geomagnetic reference field: The thirteenth generation”. *Earth Planets Space* (2022) 74:11. doi:10.1186/s40623-021-01569-z
13. Woods TN, Harder JW, Kopp G, Snow M. Solar-cycle variability results from the solar radiation and climate experiment (SORCE) mission. *Sol Phys* (2022) 297:43. doi:10.1007/s11207-022-01980-z
14. [Dataset] ISGI. ISGI, international service for geomagnetic indices (2022). Isgi - international service of geomagnetic indices accessible from: [https://isgi.unistra.fr/whats\\_isgi.php](https://isgi.unistra.fr/whats_isgi.php) (Accessed Oct 21, 2022).
15. [Dataset] British Geological Survey. British geological survey 1998 - 2017 (c) NERC (2017). British geological survey data services home page accessible from: [https://geomag.bgs.ac.uk/data\\_service/home.html](https://geomag.bgs.ac.uk/data_service/home.html) (Accessed Oct 21, 2022).
16. Gjerloev JW. The supermag data processing technique. *J Geophys Res* (2012) 117. doi:10.1029/2012JA017683
17. Coster A, Rideout WE. Utilizing the madrigal database in the whpi initiative. In: *AGU fall meeting abstracts*, Vol. 2019 (2019). p. SH43A–03.
18. Knuth J, Lucas G, Pankratz C, Berger T. The SWx TREC space weather data portal: Bringing data from diverse sources to the community. In: *AGU fall meeting abstracts*, Vol. 2021 (2021). p. SM52A–08.
19. [Dataset] SWx TREC. *Space weather data portal* (2022). Available from: <https://lasp.colorado.edu/space-weather-portal/home> (Accessed October 27, 2022).
20. Borgogno O, Colangelo G. Data sharing and interoperability: Fostering innovation and competition through APIs. *Comput L Security Rev* (2019) 35:105314. doi:10.1016/j.clsr.2019.03.008
21. Weigel RS, Vandegriff J, Faden J, King T, Roberts DA, Harris B, et al. Hapi: An api standard for accessing heliophysics time series data. *JGR Space Phys* (2021) 126:e2021JA029534. doi:10.1029/2021JA029534
22. Angelopoulos V, Cruce P, Drozdov A, Grimes EW, Hatzigeorgiou N, King DA, et al. The space physics environment data analysis system (SPEDAS). *Space Sci Rev* (2019) 215:9. doi:10.1007/s11214-018-0576-4
23. Shumko M, Chaddock D, Gallardo-Lacourt B, Donovan E, Spanswick EL, Halford AJ, et al. AuroraX, PyAuroraX, and aurora-asi-lib: A user-friendly auroral all-sky imager analysis framework. *Front Astron Space Sci* (2022) 9. doi:10.3389/fspas.2022.1009450
24. Hutko AR, Bahavar M, Trabandt C, Weekly RT, Fossen MV, Ahern T. Data products at the IRIS-DMC: Growth and usage. *Seismological Res Lett* (2017) 88:892–903. doi:10.1785/0220160190
25. Hosseini K, Matthews KJ, Sigloch K, Shephard GE, Domeier M, Tsekhmistrenko M. Submachine: Web-based tools for exploring seismic tomography and other models of Earth's deep interior. *Geochem Geophys Geosyst* (2018) 19:1464–83. doi:10.1029/2018GC007431
26. Ireland J, Hughitt K, Müller D, Dimitoglou G, Schmiedel P, Fleck B. The helioviewer project: Discovery for everyone everywhere. In: *AAS/Solar physics division meeting# 40*, Vol. 40 (2009). p. 15–01.
27. [Dataset] Grayzeck E, Bell E, Hills K. *National space science data center* (2014).
28. Gillies DM, Donovan E, Hampton D, Liang J, Connors M, Nishimura Y, et al. First observations from the trex spectrograph: The optical spectrum of steve and the picket fence phenomena. *Geophys Res Lett* (2019) 46:7207–13. doi:10.1029/2019GL083272
29. Barry KM, Cavers DA, Kneale CW. Recommended standards for digital tape formats. *GEOPHYSICS* (1975) 40:344–52. doi:10.1190/1.1440530
30. Jeiran M, Vogel BI, Miller KJ. Common data format. In: K Jaskie, TL Overman, RI Hammoud, A Mahalanobis, editors. *Automatic target recognition XXXII*. Bellingham, WA, USA: SPIE (2022). doi:10.1117/12.2618565
31. Wang S, Wang J, Zhan Q, Zhang L, Yao X, Li G. A unified representation method for interdisciplinary spatial Earth data. *Big Earth Data* (2022) 2022:1–20. doi:10.1080/20964471.2022.2091310
32. Pfander I, Johnson H, Arms S. Comparing read times of zarr, HDF5 and netCDF data formats. In: *AGU fall meeting abstracts*, Vol. 2021 (2021). p. IN15A–08.
33. Weber T, Kranzlmüller D. How FAIR can you get? Image retrieval as a use case to calculate FAIR metrics. In: 2018 IEEE 14th International Conference on e-Science (e-Science). Amsterdam, Netherlands: IEEE (2018). doi:10.1109/escience.2018.00027
34. Palma SD, Nucci DD, Palomba F, Tamburri DA. Toward a catalog of software quality metrics for infrastructure code. *J Syst Softw* (2020) 170:110726. doi:10.1016/j.jss.2020.110726
35. Lowenberg D. Recognizing our collective responsibility in the prioritization of open data metrics. *Harv Data Sci Rev* (2022) 4. doi:10.1162/99608f92.c71c3479
36. Jonas S, Fronczyk K, Pratt LM. A framework to understand extreme space weather event probability. *Risk Anal* (2018) 38:1534–40. doi:10.1111/risa.12981
37. Robinson R, Zhang Y, Garcia-Sage K, Fang X, Verkhoglyadova OP, Ngwira C, et al. Space weather modeling capabilities assessment: Auroral precipitation and high-latitude ionospheric electrodynamic. *Space Weather* (2019) 17:212–5. doi:10.1029/2018SW002127
38. Camporeale E. The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather* (2019) 17:1166–207. doi:10.1029/2018SW002061
39. Gombosi TI, Chen Y, Gloer A, Huang Z, Jia X, Liemohn MW, et al. What sustained multi-disciplinary research can achieve: The space weather modeling framework. *J Space Weather Space Clim* (2021) 11:42. doi:10.1051/swsc/20211020
40. Kauristie K, Andries J, Beck P, Berdermann J, Berghmans D, Cesaroni C, et al. Space weather services for civil aviation—Challenges and solutions. *Remote Sensing* (2021) 13:3685. doi:10.3390/rs13183685
41. Morley SK. Challenges and opportunities in magnetospheric space weather prediction. *Space Weather* (2020) 18. doi:10.1029/2018sw002108
42. Licata RJ, Tobiska WK, Mehta PM. Benchmarking forecasting models for space weather drivers. *Space Weather* (2020) 18. doi:10.1029/2020sw002496
43. Angryk RA, Martens PC, Aydin B, Kempton D, Mahajan SS, Basodi S, et al. Multivariate time series dataset for space weather data analytics. *Sci Data* (2020) 7:227. doi:10.1038/s41597-020-0548-x
44. Tobiska WK, Bowman BR, Bouwer SD, Cruz A, Wahl K, Pilinski MD, et al. The SET HASDM density database. *Space Weather* (2021) 19. doi:10.1029/2020sw002682
45. Pandey C, Ji A, Angryk RA, Georgoulis MK, Aydin B. Towards coupling full-disk and active region-based flare prediction for operational space weather forecasting. *Front Astron Space Sci* (2022) 9. doi:10.3389/fspas.2022.897301
46. White RL, Accomazzi A, Berriman GB, Fabbiano G, Madore BF, Mazzarella JM, et al. The high impact of astronomical data archives. In: *astro2010: The astronomy and astrophysics decadal survey*, Vol. 2010 (2009). p. P64.
47. Whitlock MC, McPeck MA, Rausher MD, Rieseberg L, Moore AJ. Data archiving. *The Am Naturalist* (2010) 175:145–6. doi:10.1086/650340
48. Piwowar HA, Vision TJ, Whitlock MC. Data archiving is a good investment. *Nature* (2011) 473:285. doi:10.1038/473285a
49. Tenopir C, Christian L, Allard S, Borycz J. Research data sharing: Practices and attitudes of geophysicists. *Earth Space Sci* (2018) 5:891–902. doi:10.1029/2018ea000461
50. Florio M. *Investing in science: Social cost-benefit analysis of research infrastructures*. Cambridge, MA, USA: MIT Press (2019).