Cover Page

Developing a Vision for Heliophysics Infrastructure

Co-authors: Rebecca Ringuette[1, 2], Ryan M. McGranaghan[2,3], Alexander Engell[4], Thomas Y. Chen[5], Barbara J. Thompson[2].
[1]ADNET Systems Inc, 6720B Rockledge Dr., Suite 504, Bethesda, MD 20817, USA,
[2]NASA Goddard Space Flight Center, Greenbelt, MD 20769, USA
[3]Orion Space Solutions (OSS), 282 Century Place, Suite 1000, Louisville, CO 80027, USA
[4]NextGen Federal Systems, 1399 Stewartstown Road, Suite 350, Morgantown, WV 26505, USA
[5]Columbia University, New York, NY 10027, USA

Synopsis: (limit of 400 characters)

The current infrastructure in heliophysics acts as a barrier towards progress due to the lack of intelligent connections between the many valuable resources in our field. Due to the large scale of this problem, the path to a productive infrastructure requires community collaboration towards a common vision. This white paper presents a broad overview of what that vision could be and a path forward.

Current Conditions:
The infrastructure of the heliophysics community consists of many valuable components, but lacks the connections needed by its members to efficiently make use of these promising resources. Researchers spend the majority of their time discovering and implementing data, models, software, and high-performance computing resources, often duplicating what others have done in the process, such as described in a recent meeting of the Python in Heliophysics Community (link below). The careers of those who have contributed to the development of these resources are often side-lined by the incomplete and often completely lacking citations to their work, especially for software (e.g. Niemeyer et al. 2021 and Ringuette et al. 2022a). Interested students and researchers in other fields are typically directed to unreliable sources for basic and intermediate topic explanations when using basic internet search interfaces. Although multi-disciplinary and cross-disciplinary research is encouraged, the supporting structure for pursuit of such projects is dismal at best, compromising and even stopping most projects before they gain momentum (e.g. Heliophysics Advisory Committee 2018). Researchers in many countries cannot access the infrastructure needed to perform research with large datasets, such as model data outputs and solar imagery, due to limited access to the required cyberinfrastructure and hardware. Isolated pockets of the community are struggling to improve the Heliophysics infrastructure, but lack coordination between them (e.g. the efforts by the Astrophysics Data System[1]. **The heliophysics community suffers universally from the gaping holes in the infrastructure, blocking our transformation to open science. Pockets of vision exist, but lack sufficient financial backing from the funding agencies to actually coalesce the existing resources and implement the proposed solutions.** As a result, we make slow incremental progress towards the infrastructure solutions needed.

Building a Vision:
No single person has the complete vision for the infrastructure the heliophysics community needs, but everyone has valuable input. So how does one build a vision of what the solution looks like? The authors conceive of two approaches:

- Appoint a team to perform an official survey of what the community needs, and how the community envisions those needs being fulfilled.
- Propose a possible solution, and collect feedback on how that solution can be improved.

The Heliophysics Infrastructure Workshop (Thomas et al. 2021) performed a sample execution of the first task, albeit only with a small sample of members of the heliophysics community. Their effort should be expanded and extended to involve all community members willing to participate. This white paper takes the second approach, while also recommending the first be pursued in parallel.

Two recently published papers, *Ringuette et al. (2022b) and Ringuette et al. (2022c), together propose a vision for the heliophysics infrastructure***,** including possible extensions to other communities to better connect with those fields. Those two papers describe an online LIbrary KnowledgE and Discovery (LIKED) resource for discovering and implementing knowledge, data, and infrastructure resources; and an online analysis ecosystem to simplify Discovery, Implementation, Analysis, Reproducibility, and Sharing (DIARieS) of scientific results and environments. Both are motivated by the question, "What infrastructure would a researcher or student new to Heliophysics, and possibly also new to programming, want to more easily brainstorm, network, and begin a new research project, even

[1] https://ui.adsabs.harvard.edu/

involving multiple disciplines?" The LIKED resource is a vision to address the current challenge of resource findability with some application for implementation. The DIARieS ecosystem builds upon the LIKED resource to approach the same question by applying recent advances in software and technology to decrease the difficulty of resource accessibility, interoperability, and reusability (FAIR: Wilkinson et al. 2016).

The problem of findability in heliophysics is much larger than data and software, which tends to be the focus of many recent discussions on the topic. LIKED proposes a solution to this larger problem by intelligently connecting the complete range of resources, including educational resources such as "What If?" interactive scenario interfaces and educational articles with verified content, research components such as data archives and software, and other related items such as legislation and people. As explained in Ringuette et al. (2022b), one main component of the LIKED resource is a library built on a more advanced version of Wikipedia to intelligently link a large array of resources together in a discipline-agnostic approach. This library offers intelligently linked content to the users centered around a given phenomenon, including verified descriptions in a range of difficulty levels, example analysis tutorials, "What If?" interactive scenarios (e.g. Fig. 6 of the paper), and links to related research components such as archives and publications. Another main piece of the library is a system of systems approach to linking archives together across disciplines. This is achieved by parameterizing the archives (instead of the datasets) by a reduced set of parameters, such as phenomena and dates, to more efficiently connect users to the archives and research components they are searching for. Another important component of the envisioned system of systems approach is to aim for resources of the same category to offer the same set of microservices, such as all data archives all offering quick-look graphics and sample read, write, and plotting scripts. Choosing these more unifying approaches to the findability challenge will result in a solution that is easily extensible to other disciplines without large changes (see Ringuette et al. 2022b for more details).

The vision of the DIARieS ecosystem addresses the accessibility, interoperability, and reusability challenges that the Heliophysics community faces once the various research components are found. DIARieS is envisioned to be an analysis ecosystem where users can implement the various research components into an online containerized environment with the click of a button. Imagine an online containerized version-controlled Jupyter notebook with the sharing and collaboration capabilities of a Google document, the widget-based interactions of an Excel spreadsheet for a variety of applications, including visualization generation and analysis, the integrated high-performance computational capabilities of AWS, the reproducibility of an executable paper, and more advanced narration capabilities than an iPoster. That is the essence of the DIARieS ecosystem (see Ringuette et al. 2022c for more details). The technologies required for this ecosystem exist today, and only need to be combined into a single interface to address the coming challenges of applying open science standards to our research workflows. Combined, the visions of the LIKED resource and the DIARieS ecosystem offer a complete solution for Heliophysics resources, but are too large for any one group to create. Agency support and funding of the creation of these resources is required to significantly advance our resource infrastructure.

Multiple applications of these new resources are described in the papers, and include education, communication with the commercial industry and policy-makers, research transparency, and improved

virtual collaboration. Implementing this vision will equalize access and discoverability for the heliophysics infrastructure resources we currently have, carry our field further to make data *and research* Findable, Accessible, Interoperable, and Reusable (FAIR, Wilkinson et al. 2016), and simplify transparency and reproducibility of our research and development results (e.g. open science, see NASA Open Science below). Together, **the proposed new infrastructure components outlined in this white paper and the associated references will close many of the current gaps in heliophysics' infrastructure, enable community members to more efficiently use the resources already present, lower the barriers to these resources for all, and increase the return on our investments**.

Other white papers also point to these gaps but with different approaches, some of which **can be combined with the solutions referenced here to make a more capable solution**. For example, McGranaghan et al. (2022) suggest a data science approach based on knowledge graphs to aid users in understanding the connections between different topics, data sets, and even individuals' efforts. However, building a knowledge graph requires a defined set of previously existing and known relations between objects in the system, which does not currently exist for the heliophysics infrastructure. Combining this approach with the LIKED library resource provides the **set of metadata relations needed** for the network of knowledge graphs required by McGranaghan et al. (2022), and will provide a powerful portal for community members and guests to gain understanding and make connections in heliophysics.

On the computational side of the issue, two noteworthy examples are the developing analysis ecosystem called the Space Radiation INtelligence System (SPRINTS, Engell et al. 2017) and the Kamodo platform (Pembroke et al. 2022). The SPRINTS ecosystem is maturing into a powerful tool for heliophysics, quite similar to many of the capabilities called for in the DIARieS paper referenced below. Similarly, the Kamodo platform is pursuing computational capabilities on the cloud with large data sets. Discussion and collaborations need to be encouraged between these and similar efforts, such as HDRL's (Heliophysics Data Resource Library) online computational platform, to accelerate our progress towards open science.

The structure behind these ideas are topic-agnostic, so they are **fully extensible to other fields**, which can lead to invaluable connections to *and guidance from* other disciplines (e.g. Earthcube[2], and Gelu et al. 2020 and 2022).

Recommendations:

These new infrastructure elements are to be built by the community, with funding, guidance, and coordination opportunities given by appointed representatives from the community. Specific examples of contributions to the DIARieS ecosystem and the LIKED resource are given in each paper referenced, with more general recommendations below.

- **Conduct an official survey** of what the community needs, and how they envision those needs being fulfilled.

[2] https://www.earthcube.org/ and https://sites.google.com/njit.edu/athdmiec-rcn/home

- **Build an online library resource** to equalize discovery and access of the resources we have invested in, such as the LIKED online resource (Ringuette et al. 2022b).
- **Develop an online analysis ecosystem** to simplify implementation of the current heliophysics infrastructure resources, such as the DIARieS ecosystem (Ringuette et al. 2022c).
- **Combine ideas** proposed in white papers, such as McGranaghan et al. (2022), and those otherwise made public to improve the design of the new resources and the impact of heliophysics resources on the community.
- **Publicize a detailed list of needed elements** to improve the impact of various infrastructure resources, preferably as funded opportunities.
- **Involve members of the heliophysics community** in the development of these new infrastructure resources by hosting workshops to test the capabilities of the new components and gather feedback. This will keep the infrastructure development aligned with the needs of the community, and improve the functionality of the new components.
- **Collaborate with commercial, government, and academic** resources to more accelerate our progress towards open science.
- **Fund** this effort on a level **equivalent to a long-term mission**, but distribute the funds to the groups and companies contributing to the solution via competition.
- **Work with leaders in other communities** to gain external perspectives on our infrastructure and approaches those fields used that can be helpful for heliophysics.

The heliophysics infrastructure consists of many valuable components, but they remain difficult to find with problematic barriers hindering implementation. Several are working towards resolving these gaps, but the efforts remain disjointed. The community needs to agree on a vision and organize our efforts around that vision to more efficiently deliver impactful infrastructure resources to the community. Without this coordination, the gaps in the heliophysics infrastructure will remain a significant barrier to efficient research and development for a long time to come, depleting our community of new and developing talent. Just as with the development and construction of a long-term satellite mission, **we must work together to build a vision of the infrastructure that will most benefit the community, and then collaborate to construct, assemble, and test all the necessary pieces individually and as a unit.** Only then will we make significant and efficient progress towards closing the gaps in the heliophysics infrastructure and thus squeeze more science out of our dollars. ***Substantial progress towards open science will be impossible without a significant investment in our resource infrastructure immediately.***

References:

A.J. Engell, D.A. Falconer, M. Schuh, J. Loomis, D. Bissett (2017). SPRINTS: A framework for solar-driven event forecasting and research. *Space Weather*, **15**, pp. 1321-1346, https://doi.org/10.1002/2017SW001660.

Heliophysics Advisory Committee (2018). Framework for HPAC Feedback to Research and Analysis (R&A) Charge. Presented at the *Heliophysics Advisory Committee Meeting*, April 5-6, https://science.nasa.gov/science-red/s3fs-public/atoms/files/R&A%20Charge%20HPAC%20Feedback.pdf.

McGranaghan, R. M, A. Cameron, S. Klein, R. Ringuette, A. Halford, et al. (2022). The Heliophysics Knowledge Commons. Submitted to *The 2024 Heliophysics Decadal Survey*.

NASA Open Science: https://science.nasa.gov/open-science-overview.

Niemeyer, K. E., A. M. Smith, and D. S. Katz (2021). The Challenge and Promise of Software Citation for Credit, Identification, Discovery, and Reuse. *Journal of Data and Information Quality*, **7**, 4, https://doi.org/10.1145/2968452.

Nita, G., A. Ahmadzadeh, S. Criscuoli, A. Davey, D. Gary et al. (2022). Revisiting the Solar Research Cyberinfrastructure Needs: A White Paper of Findings and Recommendations. *arXiv*, https://arxiv.org/abs/2203.09544.

Nita, G., M. Georgoulis, I. Kitiashvili, V. Sadykov, E. Camporeale et al. (2020). Machine learning in heliophysics and space weather forecasting: a white paper of findings and recommendations. *arXiv*, https://arxiv.org/abs/2006.12224.

Pembroke, A., De Zeeuw, D., Rastaetter, L., Ringuette, R., Gerland, O., et al., 2022. Kamodo: A functional api for space weather models and data. Submitted to *JOSS*. https://github.com/openjournals/joss-reviews/issues/4053.

Python in Heliophysics Community (PyHC): Presentation detailing the access problems (beginning at about the 16 minute mark), given by Laura Sandoval and Allison Liu during a PyHC telecon on September 29, 2021. Link to recording: https://o365coloradoedu-my.sharepoint.com/:v:/g/personal/juba8233_colorado_edu/EdHJW4sk4tRPr-3WkogKH48BL-b3XzlBtGtHLfD6S8qQ2w?e=wm43NR

Ringuette, R., K. Reardon, N. Murphy, M. Petrenko, J. Rigler et al. (2022a). Advocating for Equality of Contribution: The Research Software Engineer (RSE). A white paper submitted to the Decadal Survey for Solar and Space Physics (Heliophysics) 2024-2033. [redacted].

Ringuette, R., McGranaghan, R. M. and B. Thompson (2022b). The LIKED Resource - A LIbrary KnowledgE and Discovery online resource for discovering and implementing knowledge, data, and infrastructure resources. *Advances in Space Research,* under review. Temporary link: [redacted]

Ringuette, R., Engell, A., Gerland, O., McGranaghan, R., and Thompson, B. (2022c) The DIARieS Ecosystem – An ecosystem to simplify Discovery, Implementation, Analysis, Reproducibility, and Sharing of scientific results and environments. *Advances in Space Research*, in press. https://doi.org/10.1016/j.asr.2022.05.012.

Thomas, B., Candey, R., Fung, S., Ireland, J., Jian, et al. (2021). Report on a Workshop to Understand Heliophysics Research Infrastructure. In *AGU Fall Meeting Abstracts*, SH44C-01.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018. https://doi.org/10.1038/sdata.2016.18.