Towards A Heliophysics Knowledge Commons

Sara Bond[1], Baptiste Cecconi[2], Edwin Henneken[3], Ryan McGranaghan[1,*], François Scharffe[4,*], Brian Thomas[1] and Donald Winston[5]

[1]*NASA/Caltech/Jet Propulsion Laboratory, Pasadena, California*

[2]*Paris Observatory - Laboratoire d'études spatiales et d'instrumentation en astrophysique - Sorbonne Center for Artificial Intelligence*

[3]*Center for Astrophysics, Harvard & Smithsonian, Cambridge, Massachusetts*

[4]*The Data Chefs, New York, USA*

[5]*Polyneme LLC, New York, USA*

Abstract

This poster presents HelioKOS, a knowledge organization system for heliophysics. This system is an outcome of HelioKNOW, a project aiming to foster scientific discovery by building a knowledge commons for managing heliophysics metadata, and therefore enabling observational and model data integration and assimilation. The system provides a robust infrastructure for the ingestion, annotation, and analysis of scientific publications, enabling researchers to integrate and analyze data from diverse sources and disciplines. The system is centered around the Heliophysics Knowledge Graph, which serves as a central repository for all data and metadata within the system. To facilitate knowledge discovery, the system provides a range of tools and components for concept-scheme development, harmonization, and tagging, as well as tools for query registration, ground-truth development, and performance measurement. The poster presents the heliophysics knowledge organization system together with a playground to experiment with concept-scheme annotations.

Keywords

heliophysics, knowledge commons, open science, metadata, discovery, knowledge graph, data integration, scientific community

1. Introduction

The advancement of science and technology is increasingly dependent on the principles of open science, which promotes transparency, collaboration, and the sharing of knowledge. Heliophysics (the study of the solar system and its effects on our lives, climate, and space technology) is an important area of scientific research that stands to benefit significantly from open science. The solar system is a complex and dynamic environment, with phenomena such as solar flares and magnetic storms having the potential to impact our daily lives, while also affecting the Earth's climate in the long term.

Moreover, the rapid progress in commercial space technology is expected to enable human travel to the moon and Mars in the near future. A better understanding of the solar system

Posters, Demos, and Industry Tracks at ISWC 2024, November 13–15, 2024, Baltimore, USA

*Corresponding author.

sara.a.bond@jpl.nasa.gov (S. Bond); baptiste.cecconi@obspm.fr (B. Cecconi); ehenneken@cfa.harvard.edu (E. Henneken); ryan.m.mcgranaghan@jpl.nasa.gov (R. McGranaghan); francois@thedatachefs.com (F. Scharffe); brian.a.thomas@nasa.gov (B. Thomas); donny@polyneme.xyz (D. Winston)

is necessary to ensure the safety and success of these missions. There is a vast amount of scientific research being produced by observatories, labs, satellites, and probes monitoring the solar system. This research spans various disciplines, including physics, astronomy, complex systems, data science, and computer science, and is being conducted by international teams of scientists in different organizations. Thus, Heliophysics data span many orders of magnitude in space and time and their analysis occurs across scales of research from individual to national and international.

The integration of data and research results from these diverse sources is critical for advancing our understanding of the solar system. For example, the magnetosphere-ionosphere coupling phenomena are complex processes that involves the interaction of the solar wind with the Earth's magnetic field and upper atmosphere. Understanding this process is essential for understanding fundamental physical processes in the universe as well as predicting space weather events, which can have significant impacts on satellite operations, power grids, and communication systems. However, the data and research results relevant to these phenomena are spread across different scales, labs, and organizations, making integration and collaboration challenging.

To address these challenges, the Helio-KNOW project [1, 2, 3] aims to build a knowledge commons [4] for managing heliophysics metadata and fostering scientific discovery. This project is being developed as part of a multi-year effort involving several organizations, including NASA Jet Propulsion Laboratory, the Center for Astrophysics, Harvard and Smithsonian, and Paris Observatory. The goal of the project is to provide an infrastructure for the management of heliophysics knowledge at scale, enabling researchers to integrate and analyze data from diverse sources and disciplines. For this poster we focus on HelioKOS, a knowledge organization system for heliophysics.

2. System overview

The HelioKOS system is a comprehensive and modular platform designed to facilitate knowledge discovery and management in the field of heliophysics. The system provides a robust infrastructure for the ingestion and annotation of scientific publications, enabling researchers to search and analyze data from diverse sources and disciplines.

At a high level, the system is centered around the Heliophysics Knowledge Graph (HelioKG), which serves as a central repository for all data and metadata within the system. The Astrophysics Data System (ADS) [5] provides access to a vast corpus of scientific literature, which is ingested into the system via the Dagster-ADS component (see Figure 1). The ingested data is then stored in the HelioKG, providing a rich knowledge base for further analysis and processing.

To facilitate knowledge discovery, the system provides a range of tools and components for concept-scheme development, harmonization, and tagging. Based on Ontoportal [6], the Helioportal component provides access to a range of retrieval and corpus-tagging ontologies and taxonomies, for example the Unified Astronomy Thesaurus [7], which can be loaded into the system and used to develop new concept-schemes within the HelioKOS component. The harmonization component of HelioKOS enables the integration of multiple concept-schemes, providing a unified view of the data and enabling more effective analysis.

The system also provides tools for query registration, ground-truth development, and perfor-

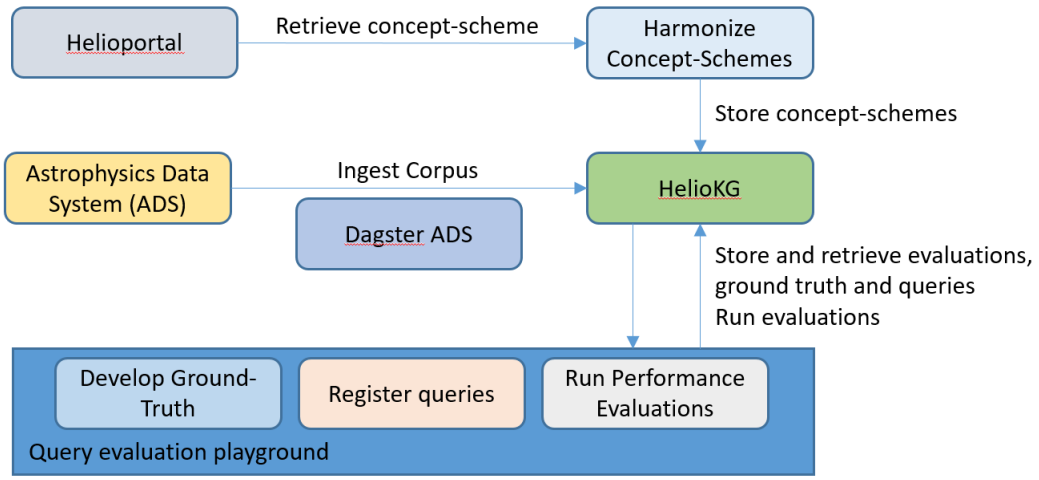

Figure 1: HelioKOS system overview

mance measurement. These tools are available under the ADS Query Eval library. The query evaluation component enables users to register queries within the system, while the ground-truth component facilitates the development of a corpus-query ground-truth in consultation with a user. The harmonization strategy component enables the registration of harmonization-application strategies, while the performance measure component enables the registration of performance measures. The performance measure job component runs performance measurements by comparing query results with and without annotations, as well as between different harmonization strategies, and stores the results back into the HelioKG. Combined together, these components allow to evaluate the effects of tagging a corpus of scientific research with a concept-scheme. They also allow to measure the effectiveness of the harmonization process over two concept-schemes.

3. Conclusion and Future Work

Overall, the HelioKOS system provides a powerful and flexible platform for knowledge discovery and management in the field of heliophysics. By integrating data from diverse sources and disciplines, the system enables researchers to gain new insights and make more informed decisions, ultimately advancing our understanding of the solar system and its impact on our lives.

Our future work as part of HelioKNOW includes providing a set of guidelines for publishing and indexing heliophysics resources on the Web. We believe that these guidelines will promote data sharing and interoperability, enabling researchers to integrate and analyze data from diverse sources and disciplines. Additionally, we are developing an ontology of heliophysics phenomena that will enable the enrichment of HelioKG with scientific datasets.

We hope for this poster to provide a valuable overview of our project and will generate insightful feedback from the community. We welcome suggestions and collaborations to help

guide our next steps and advance the state of the art in heliophysics knowledge management.

References

[1] R. McGranaghan, A. Halford, A. Narock, R. Ringuette, C. Bard, B. Thompson, E. Young, S. Klein, A. Cameron, S. Schonfeld, A. Higginson, The need for a Solar and Space Physics Knowledge Commons, in: Bulletin of the American Astronomical Society, volume 55, 2023, p. 276. doi:10.3847/25c2cfeb.a4634da3.

[2] R. M. McGranaghan, B. A. Thomas, B. J. Thompson, R. A. Ringuette, E. Young, S. Klein, A. Cameron, A. Narock, C. Bard, A. J. Halford, C. Shimizu, D. Winston, E. A. Henneken, Infrastructure for 21st century science: Towards a Heliophysics knowledge commons, in: AGU Fall Meeting Abstracts, volume 2022, 2022, pp. SH56A–02.

[3] R. M. McGranaghan, Complexity heliophysics: A lived and living history of systems and complexity science in heliophysics, 2024. URL: https://arxiv.org/abs/2307.03287. arXiv:2307.03287.

[4] R. McGranaghan, S. Klein, A. Cameron, E. Young, S. Schonfeld, A. Higginson, R. Ringuette, A. Halford, C. Bard, A. Narock, B. Thompson, The need for a Space Data Knowledge Commons, Structuring Collective Knowledge (2021). Https://knowledgestructure.pub-pub.org/pub/space-knowledge-commons.

[5] E. A. Henneken, M. J. Kurtz, A. Accomazzi, The ADS in the Information Age - Impact on Discovery, in: Organizations, volume 1, 2012, pp. 253–263. doi:10.48550/arXiv.1106.5644.

[6] C. Jonquet, J. Graybeal, S. Bouazzouni, M. Dorf, N. Fiore, X. Kechagioglou, T. Redmond, I. Rosati, A. Skrenchuk, J. L. Vendetti, M. Musen, Ontology repositories and semantic artefact catalogues with the ontoportal technology, in: T. R. Payne, V. Presutti, G. Qi, M. Poveda-Villalón, G. Stoilos, L. Hollink, Z. Kaoudi, G. Cheng, J. Li (Eds.), The Semantic Web – ISWC 2023, Springer Nature Switzerland, Cham, 2023, pp. 38–58.

[7] K. Frey, A. Accomazzi, The Unified Astronomy Thesaurus: Semantic Metadata for Astronomy and Astrophysics, 236 (2018) 24. doi:10.3847/1538-4365/aab760. arXiv:1801.01021.