

Active Data Product Archive Tracking (ADAPT) and the Future of Heliophysics Data Access

*Lee Frost Bargatze¹

1. Earth, Planetary, and Space Sciences, Institute of Geophysics and Planetary Physics, University of California Los Angeles

The Heliophysics Science Division of NASA sponsors research focused on the study of the dynamics of the Sun, the solar wind, and the interaction between the solar wind, the planets, and other solar wind bodies. These studies involve the use of data from a vast armada of spacecraft, sounding rockets, ground station observatories, together called the "Heliophysics System Observatory", coupled with information from data models and data simulations. Access to these data is enabled through the implementation of the NASA Heliophysics Science Data Management Policy, which stresses the need to maximize open access to scientifically relevant data products. The Heliophysics Data Environment, HPDE, has worked to support this policy through the creation of Virtual Observatories, VxOs, such as the Virtual Space Physics Observatory, VSPO, located at Goddard Space Flight Center, GSFC. The VxOs are charged with maintaining registries containing the information required to access Heliophysics data products. And, the HPDE has funded an effort to develop an XML schema called Space Physics Archive, Search, and Extract, SPASE, in order to document data product characteristics and to enable uniform data access. SPASE is the designated standard for documenting Heliophysics data products, as adopted by the Heliophysics Data and Model Consortium. SPASE has also been confirmed as the preferred worldwide standard for documenting Heliophysics data products after the approval of an IAGA Resolution submitted at the COSPAR 2018 meeting in Pasadena. The chore of producing SPASE metadata has proven to be an arduous effort. The VSPO metadata registries do not have a complete set of SPASE metadata to access the data from legacy Heliophysics missions. And, the SPASE registry content needs to be constantly updated in step with the release of data from new missions. Active Data Archive Product Tracking, ADAPT, which is a collection of UNIX shell scripts and IDL software, enables automatic generation of SPASE documents. ADAPT has been designed to use any available self-describing metadata associated with scientific data to produce XML metadata descriptions of any schema in a consistent and organized fashion. The goal of ADAPT focuses on providing blanket access to the full complement of data files stored on a targeted web server. The ADAPT effort is committed to providing accurate, precise, and complete SPASE documentation for data products through the implementation of quality control methods and the adoption of and adherence to uniform metadata populating standards. To date, ADAPT has been utilized to describe the Heliophysics data products that are stored by using the Common Data File, CDF, format served out by either the NASA Space Physics Data Facility, SPDF, Coordinated Data Analysis Web, CDAWeb, web site or the Heliophysics Data Portal, HDP. These CDF data repositories and the data portals, which are hosted at GSFC, are the primary means of community access to NASA Heliophysics data. The SPASE metadata registries are stored via the use of a Git distributed version-control system at <https://github.com/hpde>. In the spirit of OPEN Science and the initiative to pursue a future where open access to Heliophysics data products and interoperability, we strongly favor the stance that metadata content needs to be openly shared. Accordingly, the HPDE metadata are freely available upon request. This open metadata scenario has many advantages as a coordinated worldwide effort would help to reduce duplication of effort and improve the overall quality of the back-end metadata. As a consequence, the overall ability of the Heliophysics user community to discover, access, and archive data would be enhanced.

Keywords: Open Science, Informatics, Heliophysics Data Environment (HPDE), Virtual Observatories, Space Physics Archive, Search, and Extract (SPASE), Active Data Product Archive Tracking (ADAPT)