Tradeoffs in the Development of the SPASE Data Model

James R. Thieman1, Todd King2, and D. Aaron Roberts3

1National Space Science Data Center, NASA/Goddard Space Flight Center
2Institute of Geophysics and Planetary Physics, UCLA
3Heliophysics Laboratory, NASA/Goddard Space Flight Center

ABSTRACT

Topic: Easing data search and retrieval across the Space and Solar Physics data environment

SPASE is the Space Physics Archive Search and Extract project. This project is intended to provide a common base for finding and retrieving data of interest for the solar and space physics community across multiple space and solar physics archives and data centers. Common terminology that spans much of the disparate metadata being used by these data archives across the world enables data searches across multiple archives. Data retrieval is based on using a common data model for describing archives and data sets. The metadata can describe data sets. The metadata can describe

What is SPASE?

Space Physics Archive Search & Extract

An international, community-based standards organization with the goals of:

- Easing data search and retrieval across the Space and Solar Physics data environment
- Defining and maintaining a Data Model for Space and Solar Physics interoperability
- Demonstrating the Model’s viability

DATA ENVIRONMENT

TRADEOFFS

Finding Data
- Description only for “Full Dataset” Level
- Automated Metadata Mapping
- Cross-Discipline Keywording
- Conceptual Metadata
- Describing Objects

Using Data
- Description to “Granule” Level
- Manual Metadata Mapping
- Intra-Discipline Keywording
- Structural Metadata
- Describing Bytes

Finding Data vs. Using Data
- Should the SPASE Data Model support just finding data or using data as well?
- If using data is desired, is this just for data documentation or for data services?
- Will “finding” vs “using” be different in different disciplines?

Automated vs. Manual Metadata Mapping
- Automated metadata mapping from native descriptions to SPASE descriptions is quickly complicated by increasing level of detail
- What number of datasets to be mapped make it worthwhile developing an automated mapping tool?
- Tools developed for mapping metadata so far indicate a mixed automated/manual approach is best

Conceptual vs. Structural Approaches
- “Conceptual” metadata is usually defined by variable physical conditions or generic terms such as the location of a “bow shock”
- “Structural” metadata is usually defined in a “fixed” or “specific” way such as a geometrical definition of a boundary
- Conceptual metadata can result in confusing overlapping of keywords, especially across disciplines, but usually is better for a specific discipline

Summary
- These tradeoffs profoundly influence the evolution of the data model and the SPASE effort
- The community, in describing datasets, and using SPASE-based searches provides the feedback for choices among the tradeoffs
- We appreciate any level of commitment to this effort that you care to make