Toward Interoperability for Space Physics Data: The Problem and an Approach to the Solution

Ray Walker and the SPASE Consortium

Space Physics Data Analysis in the 21st Century

- We are entering an era with unprecedented data management challenges.
  - Much more sophisticated instruments are being proposed.
  - Data volumes are growing exponentially. (Future missions will produce ~ 10^{15} B of data!)
- Researchers will need data to able to readily find and access the required data.
- Studies will increasingly require data from more than one source.

Approach

- We prepared a detailed questionnaire and sent it to the project scientist or project manager for each operating solar terrestrial physics mission.

SOHO (1995)	Geotail (1992)
TRACE (1998)	Cluster (2000)
RHESSI (2002)	Polar (1996)
ACE (1997)	FAST (1996)
Wind (1994)	SAMPEX (1990)
Ulysses (1990)	IMAGE (2000)
Voyager (1977)	TIMED (2001)

Responses received as of January 2, 2003

- Supplemental information was obtained from a recent study of SEC operating missions by Bob McGuire

Data Accessibility

Mission	Validation Period (Months)	Access	Online	Online Request	Percent Available
SOHO	< 3	Unlimited	Yes	Yes	75-100
RHESSI	< 3	Unlimited[3]	Yes	Yes	75-100
ACE	3-6	Unlimited	Yes	Yes	75-100
Polar	9-12[1]	Unlimited	Yes	Yes	> 100
FAST	< 3	Unlimited	Yes	Yes	75-100
SAMPEX	< 3	Unlimited[2]	Yes[5]	No	75-100
TIMED	6-9	Unlimited	Yes	Yes	25-50
TRACE	< 3	Unlimited	Yes	Yes	75-100
Ulysses	< 3	Unlimited[4]	Yes	Yes	75-100
IMAGE	< 3	Unlimited[3]	Yes	Yes	75-100
Wind	< 3	Unlimited	Yes	Yes	75-100

1. Some instruments less
2. Password needed
3. Raw data – project only

4. High resolution data –project only
5. Not WWW accessible

Data Accessibility

- Searched for data by using the WWW- found most but not all of the sources.
- Tested each system by requesting recent data.
  - *SOHO*: Unclear interface, queries took a long time.
  - *RHESSI*: Primitive interface but located and extracted data quickly.
  - *ACE*: Found the data quickly and downloaded it.
  - *Polar*: Simple interface, located data easily and downloaded it.
  - *FAST*: Found the data quickly and downloaded it.
  - *TIMED*: Found the data quickly and downloaded it. Required registration to use.
  - *TRACE*: Nice interface, found data quickly.
  - *Ulysses*: Primitive interface, located data and downloaded it.
  - *IMAGE:* Interface required retyping time interval for each query, found images and downloaded them, had problems with digital data in CDF.
  - *Wind:* Located data through CDAW-Web and downloaded it.

Documentation

	Spacecraft			Instrument			Data Set Overview			Data Processing			Calibration			Quality		
Spacecraft	R	O	V	R	O	V	R	O	V	R	O	V	R	O	V	R	O	V
SOHO	Y	Y	Y	Y	Y	Y	N	-	-	Y	Y	Y	N	-	-	N	-	-
RHESSI	?	?	?	Y	Y	?	N	-	-	Y	Y	?	Y	Y	?	Y	Y	?
ACE	Y	Y	?	Y	Y	?	Y	Y	Y	Y	Y	?	N	-	-	Y	Y	?
Polar	Y	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y	N	Y
FAST	Y	Y	Y	Y	Y	?	N	-	-	N	Y	-	N	N	?	N	N	?
SAMPEX	?	?	?	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	-	-	Y	Y	Y
TIMED	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	-	-	Y	Y	Y	Y	Y	Y
TRACE	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Ulysses	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
IMAGE	Y	Y	Y	Y	Y	Y	N	-	-	Y	Y	Y	Y	Y	Y	N	-	-
Wind	Y	N	Y	Y	N	Y	N	-	-	Y	N	Y	N	-	-	Y	Y	Y

Key   Y-Yes

R-Required   N-No

O-Online   ?-No response received

V-Verified   --NA

Format/File Structure

Spacecraft	Documentation			Formats Used
	R	O	V	
SOHO	Y	Y	Y	FITS, CDF, SFDU, ASCII
RHESSI	Y	Y	Y	FITS, ASCII, Tables
ACE	Y	Y	?	HDF, Tables
Polar	Y	N	Y	CDF, ASCII, Tables, Other
FAST	Y	N	Y	CDF
SAMPEX	Y	Y	Y	Other
TIMED	Y	Y	Y	Other
TRACE	?	?	?	FITS, Other, SFDU
Ulysses	Y	Y	Y	CDF, ASCII, Tables, Other
IMAGE	Y	Y	Y	CDF, UDF, images
Wind	Y	Y	Y	ASCII, CDF

Where are we?

- In general the data are available to the science community in a timely manner.
  - Scientists can find the needed data provided they know where to look.
  - Much more care is being taken to document the data adequately.
  - Care is being taken to make sure that the data are preserved.
- Almost all of the data can be accessed in some form.

We are approaching an open data environment.

Where are we?

- Comparative research within missions isn't as easy as it could be and comparative research between missions is still very difficult.
  - Locating data if you don't know where to go is difficult.
  - Once a researcher locates the required data it comes in multiple formats.
  - There are no commercially or publicly available tools that allow researchers to use all of the formats currently used in space physics.
  - There are no space physics wide metadata standards (the existence of which could significantly lessen the burden of the multiple formats).

SPASE: A Start Toward Interoperability Between Independent Data Systems

- SPASE will be based on a series of fundamental concepts that can be used access and use space physics data.
- The concepts form a data dictionary that can be used to convey queries and responses between a user and various data systems.
- In practice, a user will formulate a query by using the data access system with which they are most familiar.
- The user interface will then conceptualize the user's inputs using the SPASE concepts (that is it will translate the query into an "interlingua" or intermediate language by using the SPASE data dictionary).
- The request will then be presented to each data system of interest.

SPASE: A Start Toward Interoperability Between Independent Data Systems Continued

- The other data systems will then localize the concepts to match their data model. Within each data system the request is resolved (i.e. translated from the SPASE interlingua into a query they understand).
- The results are then conceptualized and returned to the user.
- By using the SPASE interlingua an n^2 problem can be reduced to a 2n-1 problem.

The Spase Approach