



Heliophysics Legacy Data Restoration

Poster: SM33A-2498

R.M. Candey¹, E.V. Bell⁵, D. Bilitza², R. Chimiak³, J. Cooper¹, L. Garcia⁴, C. Gladney⁵, E. Grayzeck⁶, B. Harris³, H. K. Hills⁵, R. Johnson⁵, T. Kovalick⁵, N. Lal¹, H. Leckner⁵, M. Liu⁵, P. W. McCaslin⁵, R. McGuire¹, N. Papitashvili⁵, S. Rhodes⁵, A. Roberts¹, R. Yurow⁵

¹Code 670/NASA Goddard, ²George Mason University/NASA Goddard, ³Code 580/NASA Goddard, ⁴Wyle/NASA Goddard, ⁵ADNET/NASA Goddard, ⁶Code 690.1/NASA Goddard

Joint work by the NASA Space Science Data Coordinated Archive (NSSDCA) and the Space Physics Data Facility (SPDF) to upgrade datasets from older NASA missions on old tapes and move to online storage

Lessons Learned

- Data restoration is difficult and painful
- Avoid by ensuring new datasets are in good shape up front and use standard file formats and metadata
- Restored data in modern formats not only more accessible but enables future migration to new formats
- Although many datasets now online in formats that are readily analyzed, others will still require deciphering to puzzle out the data values and scientific meaning
- Put all data online (and make offsite copies), even though incomplete and poorly-documented, so any interested researchers can access
- Community interest helps prioritize restoration activities
- SPDF has accumulated expertise and tools for restoring datasets and will help interested scientists

Historical Challenges

- Data not readily accessible when archived on 7- and 9-track tapes, microfilm and microfiche and other media
- Instrument scientists held onto their own data and often only ones to understand them
- No open data policy or high level requirement to archive best data
- Obscure computer numerical and file formats
- Poor or missing documentation
- Undocumented or undecipherable formats and meaning of columns
- Data corruption and need for media refresh
- Analog media (plots and tables on microfilm)
- Need to assess if science importance has been superseded by newer instruments and datasets

Advantages of Old Datasets

- Valuable science is still buried within these datasets, since storage and processing power have dramatically increased, and when analyzed in conjunction with other data and models
- Provide long time history for trends and space climate
- Unique missions by orbit location and instrument complement
- Only now feasible and cost effective to restore some datasets depending on their scientific interest and usefulness

SPDF Holdings Today

- Over 2000 datasets, 54 million files, 130 TB
- Heliospheric, 1 AU interplanetary, Earth's Magnetosphere, Geosynchronous, Ionosphere, Thermosphere, Mesosphere, and some ground-based
- Satellite orbits and regions <<http://sscweb.gsfc.nasa.gov/>>

Missions at SPDF

1963-038C	DMSP (many)	Interball Aurora, Tail	RELAY 1, 2	SOLRAD 8, 9
ACE	DSCOVR	International Space Station ISS	ROCSAT	ST5
AE B-E	DSP	Ionosphere Sounding Satellite 2	Russian MSU Cosmos 900	STELAB
AEROS a, b	Equator-s	ISEE 1-3	Russian MSU Granat	STEREO Ahead, Behind
Alouette	ERS 13, 17	ISIS	Russian MSU Luna 17, 19, 21, 22	STP 78, 81
AMPTE	ESS/GEOS2	IUE	Russian MSU Mars 2, 4, 5, 7	TELSTAR 1, 2
Apollo 11, 12, 14, 15, 16, 17	ESRO 4	LANL (many)	Russian MSU Phobos 2	THEMIS (all)
ARCAD	FAST	MAGSAT	Russian MSU Vega-1, 2	TIMED
ARIEL 11, 13, 14	Finnish Meteor	MAP	Russian MSU Venera 2-11,13,14,16	TIP
ASTP	Galileo	Mariner 4, 6, 7, 9, 10	Russian MSU Zond 1, 3	Transit 5e5
ATS 1, 2, 5, 6	Genesis	MAVEN	S15	TWINS 1, 2
AZUR	Geotail	MESENGER	S3	Ulysses
BARREL	GOES (most)	MMS (all)	S-Cubed-A	Vanguard 3
CANOPUS	GPS	NOAA (most)	SAMPEX	VELA
Cassini	Hawkeye	OGO 1-6	San Marco	Viking
Cluster (all)	HELIOS 1, 2	OSO 1-8	SESAME	Voyager 1, 2
CNOFS	HEOS 1, 2	OV	SHIMMER	Wind
COSMOS 49, 900, 1809	Hinotori	Pioneer 6-11	Skylab	Other Explorers 4, 7, 12, 14, 15, 22, 26, 31, 32, 33, 35
CRRES	IBEX	Polar	Solar Maximum Mission	
CSSWE 11, 12	IMAGE	Prognoz 6, 7, 9, 10		
DARN	IMP 1-8	RADSAT		
DE 1, 2	INJUN 1, 3, 4, 5	Van Allen Probes (RBSP) (all)		

History of Many Spacecraft Data

- Record original spacecraft telemetry onto analog or digital 7- or 9- track tapes
- Separate into instrument files (often on old IBM, Univac, CDC, etc. computers)
- Process data, add ephemeris, and write to other tapes and microfilm (Instrument teams)
- Make a copy for NSSDC (main NASA archive at the time) and attach descriptive papers (teams)
- Somewhat later, copy onto new 9-track and 3480 cartridges (for redundancy), sometimes converting to more modern VMS or IBM numbers (NSSDC)
- At each step, NSSDC pre-pended some note pages into the dataset folder
- Scan dataset paper documents into DSC PDF files
- Recently, NSSDCA sent some degraded tapes to JBI (John Bordynuik Inc.) for scanning of the raw magnetic field bits as *.tap scan files [JBI process in a 2008 Storage Conference presentation <<http://storageconference.us/2008/presentations/3.Wednesday/5.Bordynuik.pdf>>]
- NSSDCA records NSSDC dataset IDs and record file attributes in attributes.xml file, and makes a tar file of each set (NSSDC AIP generation)
- NSSDCA sends files to SPDF as the new long-term active archive for NASA heliophysics data
- SPDF puts them on <<http://spdf.gsfc.nasa.gov/pub/data/>> in hopes they can be restored and archives them in multiple locations
- SPDF adds an HTML file from the NSSDCA Master Catalog <<http://nssdc.gsfc.nasa.gov/nmc/>> about the dataset and adds scanned documentation where available

Dataset Documentation

- NSSDCA Master Catalog of old missions and datasets <<http://nssdc.gsfc.nasa.gov/nmc/>>
- Bfiles text and digital files <spdf.gsfc.nasa.gov/pub/documents/old/miscellaneous_documents_from_nssdc/>
- DSC bitmap scans (PDFs) <spdf.gsfc.nasa.gov/pub/documents/old/documentation_from_nssdc/rena_med_dataset_catalogs/>
- Other NSSDC docs <<http://nssdc.gsfc.nasa.gov/nmc/trf-search.htm>>

Types of Old Files

- ASCII or EBCDIC with fixed length or variable-length records (various end-of-line or record-length markers)
- Binary with fixed length or variable-length records (various end-of-line or record-length markers)
- VMS Backup files
- VMS file and record attributes
- TAP files from JBI scans: convert with custom tape-read software
- ANSI label tapes: use ansi software
- NSSDC AIP format: use aip-tool

Common Data Format (CDF)

- Self-describing scientific platform- and discipline-independent data format
- CDF enables easy and powerful use in browse and analysis tools like CDAWeb <<http://cdaweb.gsfc.nasa.gov/>> and Autoplot <<http://autoplot.org/>>
- CDF Creation
 - Define and create the CDF skeleton CDF with metadata (by text editing or using the SKTeditor tool <<http://sscweb.gsfc.nasa.gov/sktditor/>>)
 - SPDF/ISTP/IACG Metadata Guidelines <spdf.gsfc.nasa.gov/sp_use_of_cdf.html>
- Open empty CDF and add data
 - In IDL, use IDLmakeCDF procedures to import ASCII, various binaries, IDL savesets <spdf.gsfc.nasa.gov/CDAWlib.html#IDLmakecdf>
 - Use makeCDF tool to import ASCII or binary files <spdf.gsfc.nasa.gov/makecdf.html>
 - Direct writes using CDF library in C, Fortran, Perl, Python, C#, Matlab
- Use CDF tools to translate among multiple self-describing data formats (HDF, netCDF, FITS, XML, ASCII)

Numerical Conversion Utilities

- conv_vax_unix_v4.pro in IDL <spdf.gsfc.nasa.gov/pub/software/format_conversion/conv_vax_unix_v4.pro>
- FitPnt Ruby routine for converting most everything <<http://float-formats.rubyforge.org/classes/FitPnt.html>>
- Univac format <<http://www.fourmilab.ch/documents/univac/minuszero.html>>
- Old formats <<http://nssdc.gsfc.nasa.gov/nssdc/formats/>>

Example Data Restorations

- Mona Kessel and Scott Boardsen restored Hawkeye (Explorer 52) in polar orbit with 20 Re apogee <spdf.gsfc.nasa.gov/pub/data/hawkeye/documents/archived_website/hawkeye.html>
- Bob Benson and Dieter Bilitza restored ISIS/Alouette Torsion Sounder from analog telemetry tapes <spdf.gsfc.nasa.gov/isis/isis-status.html>
- Several datasets restored as part of Foundations of Data Curation course, taught by Ruth Duer and Karen Baker at University of Illinois Urbana Champaign Graduate School of Library and Information Science (UIUC, GSLIS):
 - Colleen Fallaw created CDF file from IMP-1 hourly magnetic field data <spdf.gsfc.nasa.gov/pub/data/imp/imp1/fluxgate_magnetometer/hourly_averaged_magnetic_field/cdf/imp1_h0_fgm_20150507.cdf>
 - Nushrat Jahan Khan converted Injun-1 Airglow Photometry data (from EBCDIC and removed block characters) <spdf.gsfc.nasa.gov/pub/data/injun/injun1/auroral_airglow_photometers/>

SPDF needs the ongoing help of the heliophysics science community in identifying, restoring, documenting, and re-archiving still important legacy data

All SPDF Data and Services can be reached at spdf.gsfc.nasa.gov