Activities and Status of the Space Physics Data Facility (SPDF) Archive for the Past Year

Robert Candey, Lan Jian, and the rest of the SPDF team

Fourth IHDEA Meeting

2021 Sept. 27 – Oct. 1

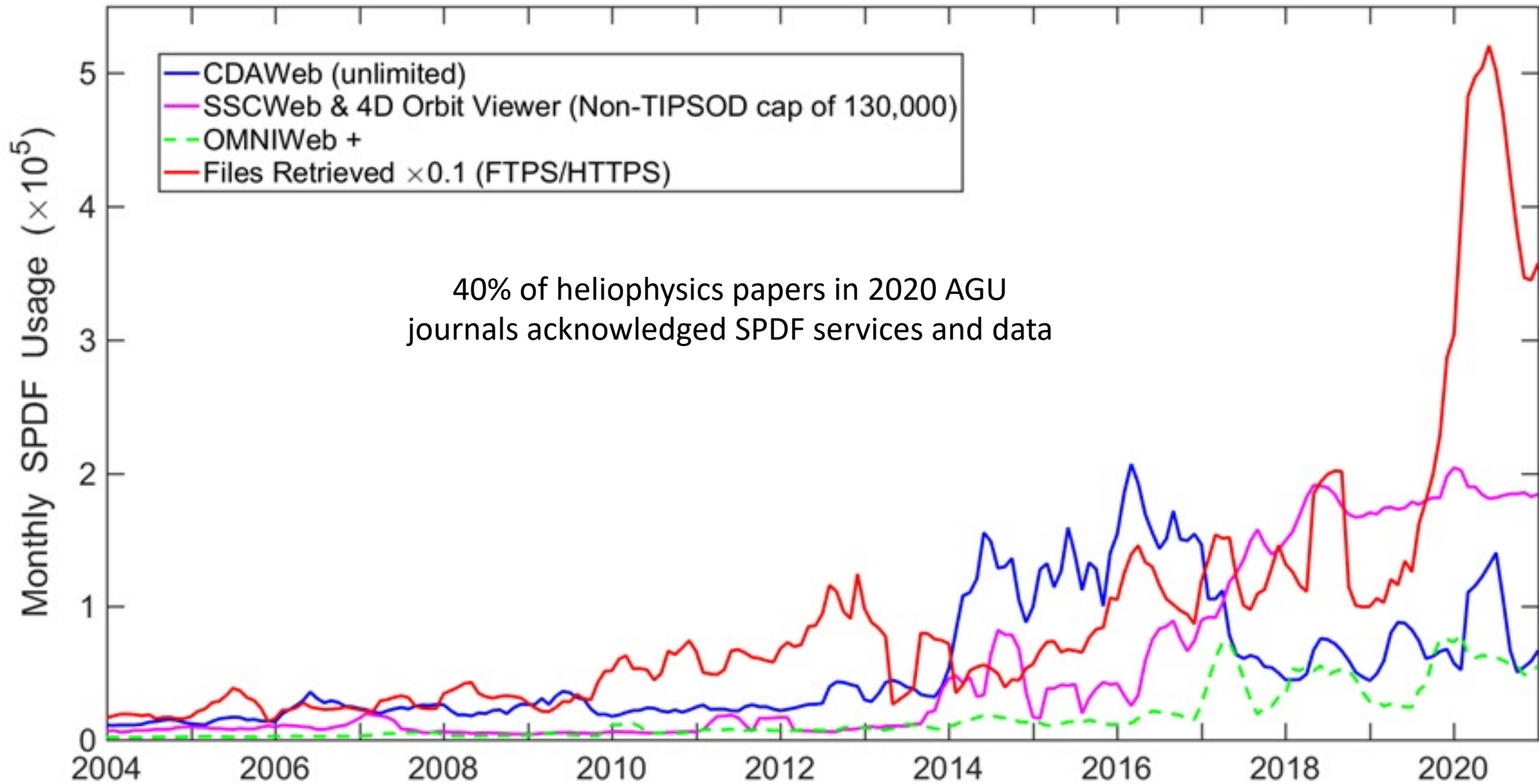SPDF <https://spdf.gsfc.nasa.gov>

SPDF activities in past year

- Added many new datasets from ICON, GOLD, Parker Solar Probe (87), IBEX (40), Solar Orbiter (81), MMS (16), BARREL (219), FAST, Voyager PWS waveform, and many other spacecraft, rocket, balloon, and ground instruments

- Final data from Van Allen Probes (RBSP), with most datasets entirely reprocessed

- Automated ingest pipeline for > 75 missions out of over 200 missions for a total of ~4,000 datasets using ~350 TB (ingest and usage logs: https://cdaweb.gsfc.nasa.gov/publiclogs/)

- Recent average monthly data ingestion rate: ~0.6 million files, ~13.7 TB data

- Creating CDFs from SOHO in-situ data and finishing making CDFs for IBEX data and Wind STICS

- Continue population of OMNI, COHO, SSC databases

- CDAWeb plot and display improvements, waveforms, inventory plots, time slices, audification

- Adding SPASE Resource IDs and DOIs to CDAWeb metadata and displays

- Working towards a grand vision as part of NASA's Heliophysics Digital Resource Library, including 6 new curation scientists (part-time)

- Recent passing of Nand Lal, and retirements of Reine Chimiak and John Cooper

Planned SPDF activities

- Standardize ISTP/IACG Metadata Guidelines with version control, etc.

- Adding some global attributes and variable attributes to ISTP metadata standard (e.g., author list for DOIs, DOI, Variable_display_order, Variable_display_indent_level, Associated_parent_variable, Dataset_group, Mission_parent) and from Cluster/Solar Orbiter: Representation, Tensor_order, Coordinate_systems, Rotation_matrices, Unit_quarternion

- New SKTeditor in Javascript or Python, including defining SPASE metadata at the same time as defining the internal metadata and structure of the CDF/netCDF dataset being created

- Working on web services for event lists for burst mode data and science events (CMEs, bow shock crossings, etc.) and use by SSCweb and CDAWeb to better serve intermittent/burst data (find next/previous burst)

- Developing Javascript alternative to the Java-based 4D Orbit Viewer

- HTML5/Javascript-based browser interface for CDAWeb/SSCweb, building on the above prototype 4D orbit viewer, expanded to add interactive data plotting and sonification tied to the orbit display, perhaps with data glyphs along the orbits as well (uses JSON output from SSCweb and CDAWeb web services)

- Quick start guides, tutorials, improved documentation, web redesign

SPDF Statistics
(see reports at <https://cdaweb.gsfc.nasa.gov/publiclogs/>)

40% of heliophysics papers in 2020 AGU
journals acknowledged SPDF services and data

SPDF action items from previous IHDEA meetings

1. Action on SPDF: to circulate on a regular basis, at least yearly, to IHDEA members the roadmap of CDF development
   - See next page for CDF status

2. Action on SPDF: ISTP rules to allow filenaming with start and end dates, in such a way that these filenames will not cause errors or warning by SKTeditor
   - In progress, mostly complete

3. Action on NASA and ESA/ESDC: to coordinate/facilitate the transfer of all science datasets of the Cluster mission from ESDC to SPDF
   - Finalizing plan for copying Cluster data, will start ingest soon

CDF Status and Recent Development

- About to release CDF 3.8.1 with Unicode UTF-8 encoding support

- Continued CDF support and general development, plus added features

- High-level functions to read variables or whole CDF into a map structure for IDL, Java, Perl and C#

- CDF-JSON converter

- Improved and signed Windows installer, notarized Mac installer, autoconf/make build/install

- Improved dynamic space management in core library

- More generalized CDF epoch data encoding and parsing functions, default encoded epoch data now ISO-8601 format

- CDF_TIME_TT2000 routines handle negative leap seconds; probably required in next 5 years

- CDF epoch data conversion to/from Unix time

- String typed variables padded with a single space and followed by NULs

- cdf2skt tool option to choose how to display variable's metadata and data

- Added CDF magic numbers to the Unix FILE utility http://www.darwinsys.com/file/

CDF Plans

- Improve documentation, beginner's guides, add to Wikipedia CDF entry

- Looking into supporting CDFs in cloud object storage, perhaps Zarr like netCDF is exploring

- Define CDF MIME type and international standard

- Apache 2 license in place of current custom license

- Update CDFML and its corresponding JSON representation with cdf.xsd to use more specific datatypes (e.g., xs:dataTime, xs:integer, xs:float, etc.) instead of just xs:string

- Add support for CDF to command line netCDF tools, such as NCO, NCAR, ANTS, NCtools

- CDF gap checker to write filename, variable name, begin and end time, number of records, and any gaps greater than a certain amount (G-good, M-missing, F-fill, R-outside range, B-backward time)

- Add CDF support to Octave, Gnu Data Language (GDL), Excel, Ruby, C++, WebWinds, LinkWinds, Opendap, SWIG.org

Backup

SPDF access protocols

- Files available through HTTPS and FTPS <https://spdf.gsfc.nasa.gov/pub/>

- HAPI <https://cdaweb.gsfc.nasa.gov/hapi>

- REST and SOAP web services for
  - CDAWeb <https://cdaweb.gsfc.nasa.gov/WebServices/>
  - SSCweb orbits <https://sscweb.gsfc.nasa.gov/WebServices/>

- Autoplot <http://autoplot.org/help#CDAWeb>

- Other methods such as IDL <https://cdaweb.gsfc.nasa.gov/alternative_access_methods.html>