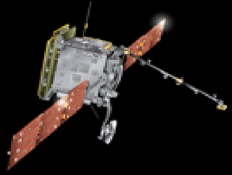
VSO 2.0 - What a new VSO means for the Heliophysics community

Alisdair Davey (adavey@nso.edu)

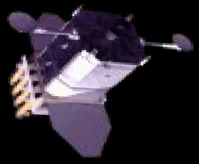
(Ashley sends their apologies!)

IDHEA Meeting, Oct 2024, Madrid, Spain

What is VSO?


Solar Orbiter

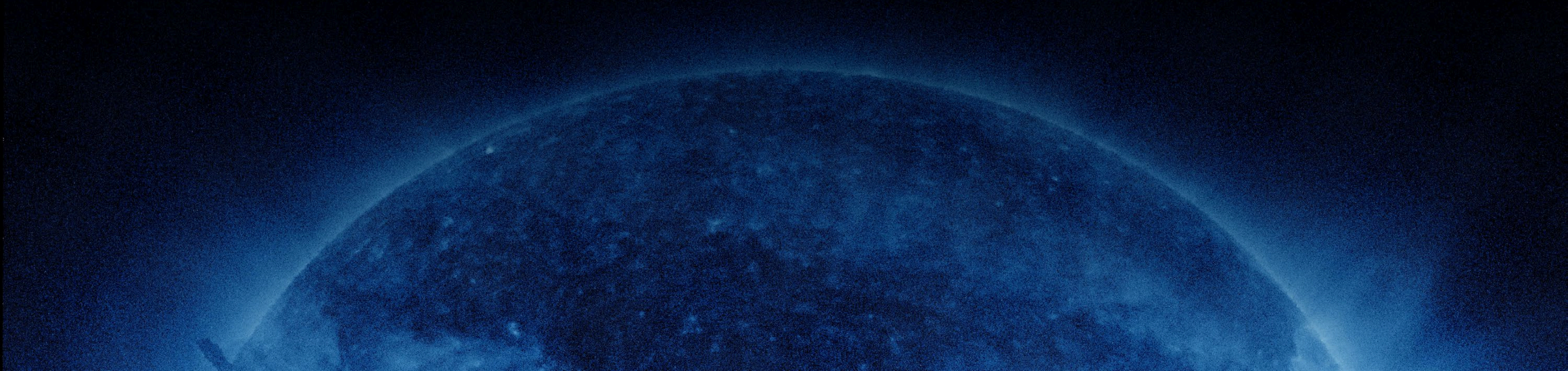
IRIS


SDO

- The Virtual Solar Observatory (VSO) is a well-established (2002) community resource that provides homogenous discovery and access to distributed Solar and Heliophysics data.
- VSO supports 25 active data providers, supporting 51 sources (satellites, rocket / balloon flights, ground-based observatories etc.) with data from 114 different instruments. Data goes back to 1869.
- VSO provides packages in the standard SolarSoft (IDL based) analysis environment and the same access to data from within SunPy (Python based). The same API is available for other languages.
  - Historically seen that used for C, Perl, Java and JavaScript.
- VSO provides a check on the quality of available data and metadata, data validation, and standards, open source and open data evangelization (SOLARNET, TOPS, FAIR etc.)

Why have we been around so long?

- A wise man once said “First, serve your community.”
  - Our quintessential “Small Box” approach.
  - Small team.
  - VSO includes “we know what you meant” technology!
- We’re part of the analysis systems (SolarSoft / SunPy).
- We’ve become glue / furniture!
- Nepalese food.

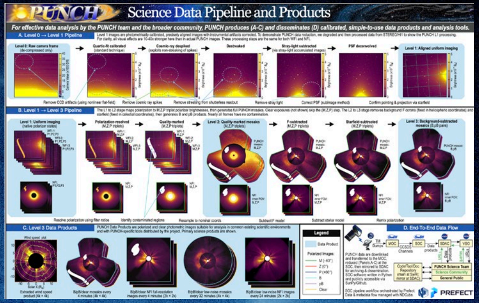


Why it's time for VSO 2.0

- VSO was first introduced in 2002. The base VSO code, including the core, registry, user interface and most data providers have been written in Perl.

Maybe the take-away is that we are modernizing VSO because we want VSO to be sustainable, and to be maintainable most probably by the next generation!

  - Lack of separation between data related and display parts.
- Our current data model makes it difficult to unambiguously categorize the data products from new missions such as PUNCH or new ground-based observatories such as DKIST.
- VSO web site is ancient! We'd really like to take advantage of over 20 years of development in web application technology!



PUNCH Science Data Pipeline and Products

For effective data analysis by the PUNCH team and the broader community, PUNCH produces (A-C) and disseminates (D) calibrated, simple-to-use data products and analysis tools.

A. Level 0 → Level 1 Pipeline

Level 1 images are photometrically calibrated, precisely aligned images with instrumental artifacts corrected. To demonstrate PUNCH data reduction, we degraded and then processed data from STEREO/HI1 to show the PUNCH L1 processing. For clarity, all visual effects are 10-40x stronger here than in actual PUNCH images. These processing steps are the same for both WFI and NFI.

B. Level 1 → Level 3 Pipeline

The L1 to L2 stage maps polarization to M,Z,P triplet polarizer brightnesses, then generates full PUNCH mosaics. Clear exposures (not shown) skip the (M,Z,P) step. The L2 to L3 stage removes background F corona (fixed in heliospheric coordinates) and starfield (fixed in celestial coordinates), then generates B and pB products. Nearly all frames have no contamination.

C. Level 3 Data Products

PUNCH Data Products are polarized and clear photometric images suitable for analysis in common existing scientific environments and with PUNCH-specific tools distributed by the project. Primary science products are shown.

D. End-To-End Data Flow

PUNCH data are downlinked and transferred to the MOC, reduced (Panels A-C) at the SOC, then mirrored to SDAC for archiving & dissemination. SOC software written in Python and publicly accessible via SunPy/Github.

SOC pipeline workflow orchestrated by Prefect. Data & metadata flow managed with NDCube.

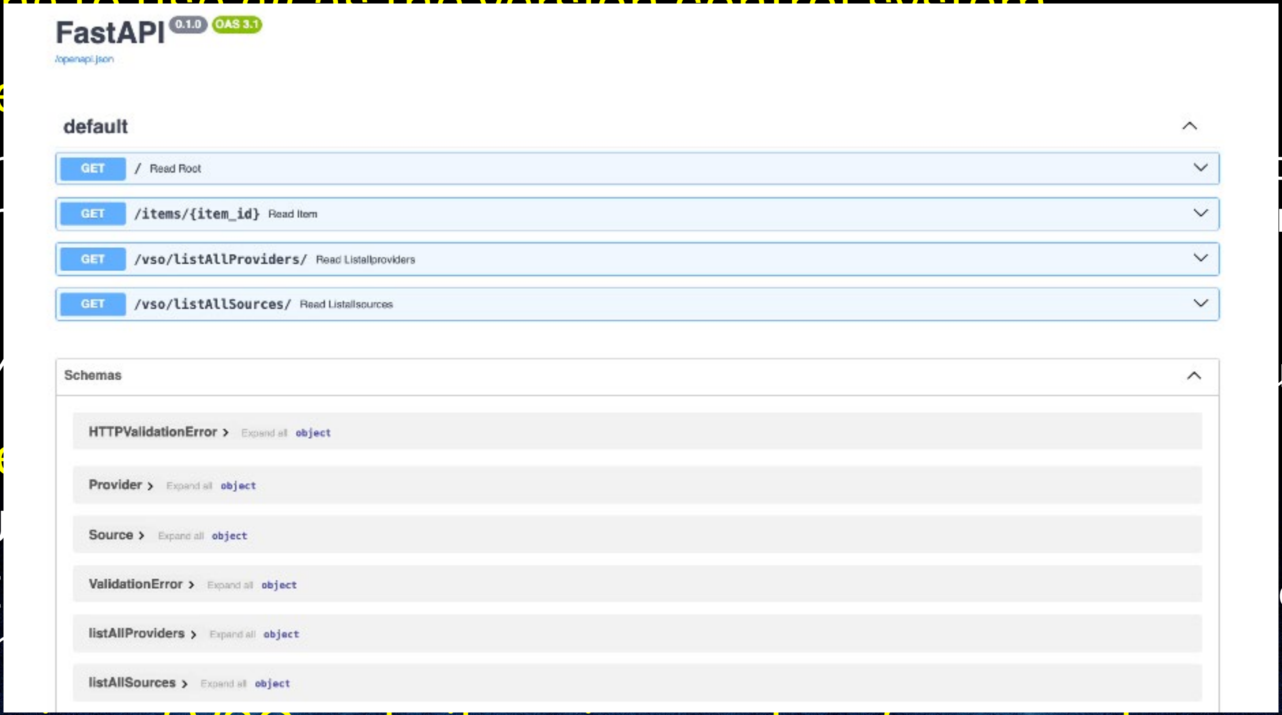
Level 0: Raw camera frame (de-compressed only)
Quartic-fit calibrated (standard technique)
Cosmic-ray despiked (exploits non-streaking of spikes)
Destreaked
Stray-light subtracted (via stray-light accumulated images)
PSF deconvolved
Level 1: Aligned uniform imaging
Remove CCD artifacts (using nonlinear flat-field)
Remove cosmic ray spikes
Remove streaking from shutterless readout
Remove stray light
Correct PSF (subimage method)
Confirm pointing & projection via starfield
Level 1: Uniform imaging (native polarizer states)
Polarization-resolved (M,Z,P triplets)
Quality-marked (M,Z,P triplets)
Level 2: Quality-marked mosaics (M,Z,P triplets)
F-subtracted (M,Z,P triplets)
Starfield-subtracted (M,Z,P triplets)
Level 3: Background-subtracted mosaics (B,pB pairs)
Resolve polarization using filter ratios
Identify contaminated regions
Resample to nominal coords
Subtract F model
Subtract stellar model
Remix polarization
Wind speed plot
Extracted wind speed product (4k × 4k)
B/pB/clear mosaics every 4 minutes (4k × 4k)
B/pB/clear NFI full-resolution images every 4 minutes (2k × 2k)
B/pB/clear low-noise mosaics every 32 minutes (4k × 4k)
B/pB/clear low-noise NFI images every 24 minutes (2k × 2k)
Legend
Data Product
Polarized Images:
M (-60°)
Z (0°)
P (+60°)
B
pB
Clear
TM Dumps
MOC
CCSDS Channels
SOC
Data products
SDAC
VSO
Code/Tool/Doc Repository (main at SwRI; mirror at SDAC)
PUNCH Science Team
Science Community
General Public
PREFECT

Farewell to camels and llamas!

4th Edition
Covers Version 5.14
Unmatched Power for Text Processing and Scripting
Programming
Perl
Tom Christiansen, br
Larry Wall & Jo
O'REILLY
Making Easy Things Easy & Hard Things Possible
Learning Perl
6th Edition
Covers Perl 5.14
O'REILLY
Randal L. Schwartz, brian d foy & Tom Phoenix

Redesigning VSO - Technical

- Moving to use Python as the main implementation language
  - Have worked closely with SunPy community over the last few years.
  - Already have data providers written in Python.
- Moving to use *git* as the version control system
- Mode
  - In ting the in n.
- Movi
  - M ures.
- Imple
  - U
  - E esign of th



FastAPI 0.1.0 OAS 3.1

/openapi.json

default

GET / Read Root

GET /items/{item_id} Read Item

GET /vso/listAllProviders/ Read Listallproviders

GET /vso/listAllSources/ Read Listallsources

Schemas

HTTPValidationError > Expand all object

Provider > Expand all object

Source > Expand all object

ValidationError > Expand all object

listAllProviders > Expand all object

listAllSources > Expand all object

- Redesign of VSO website using modern frameworks.
- *While maintaining Current VSO functionality.*

VSO is adopting a Model-View-Controller (MVC) approach

MVC is a software design pattern that separates an application's logic into three layers.

- Model

  Contains the data and data-related logic that the user interacts with, such as databases, schemas, and interfaces.

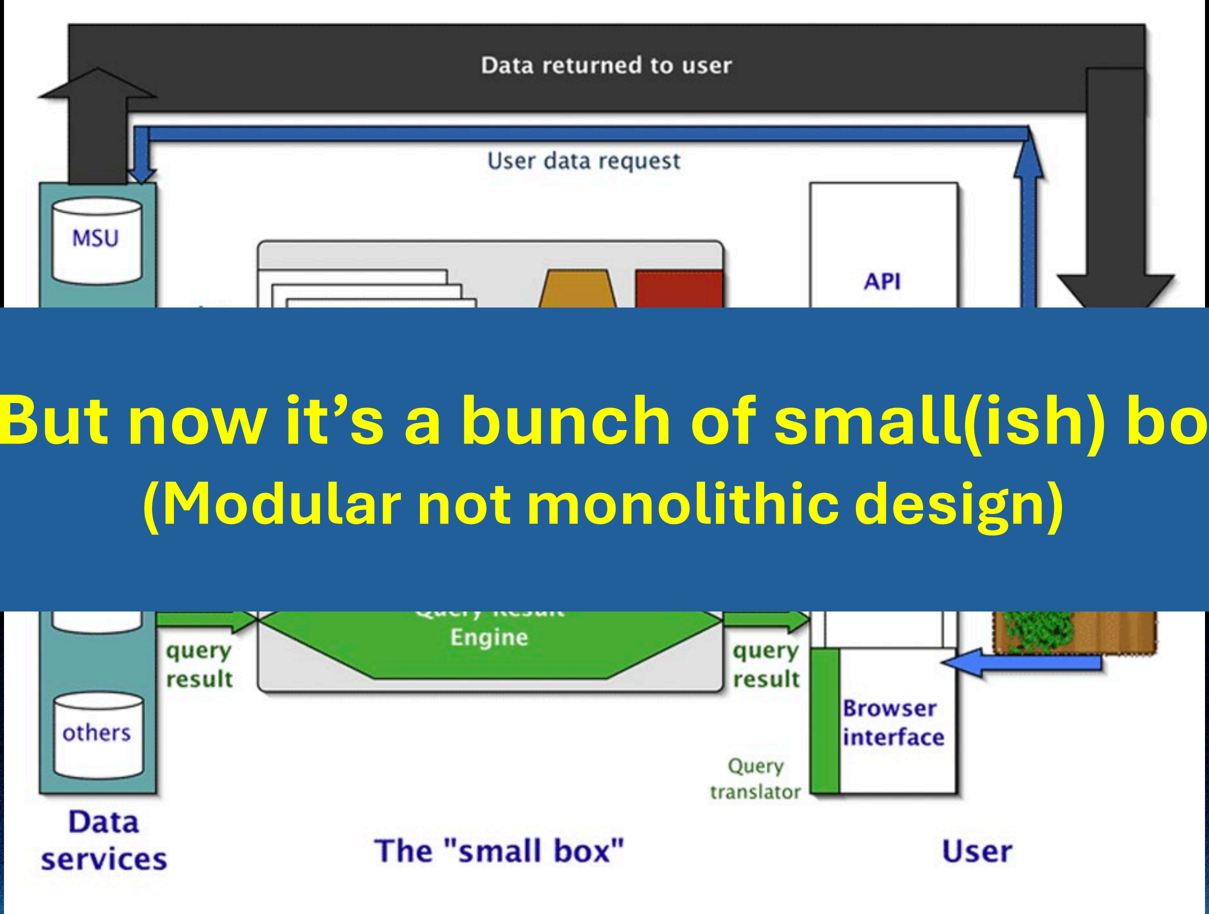
- View

  Provides the user interface (UI) that allows users to interact with the application and its data.

- Controller

  Contains the application logic that facilitates communication between the model and view and handles user interaction from the view.

The end of the small box?



No! But now it’s a bunch of small(ish) boxes!

(Modular not monolithic design)

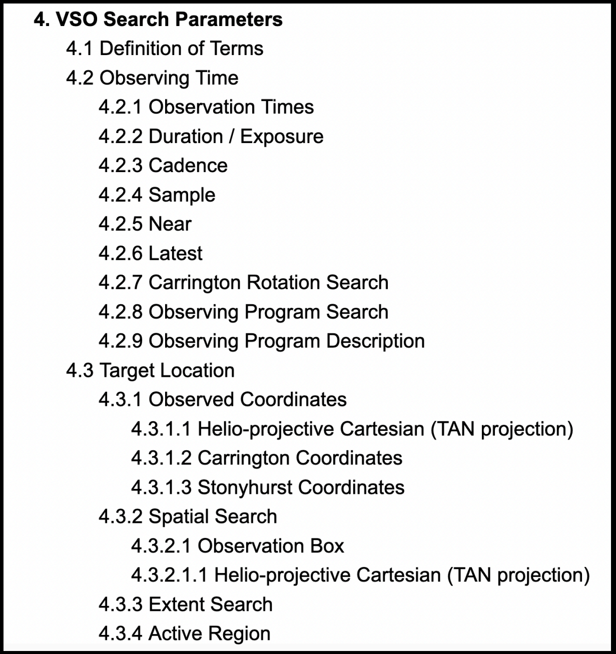
VSO 2.0 – Features

- New data model.
- Internal and external access between functional elements via a series of REST APIs.
- A new registry which uses the new data model to expose everything that each data provider can support.
- New VSO web portal and unified functionality between the portal and programmatic interfaces.
  - Saved searches. Support for user customization. Easier reproducibility!
- Updated Data Providers.
- Full spatial, and sub-second time search / support.
- Improved "health" monitoring of data providers.
- Added protocol support.
- Automated statistical tracking of queries and downloads.
- Provide more responsive services. Better error handling and propagation. Unambiguously able to say why when there are no results.
- Greatly improved documentation. (Python notebooks etc.)
- Opportunities for people outside of the core VSO team to contribute.

PUNCH

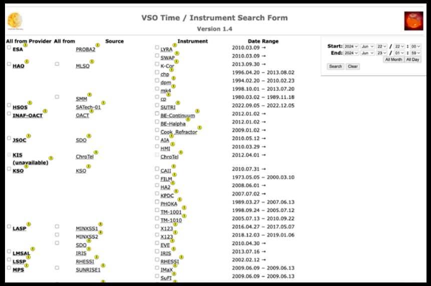
VSO Data Model 2.0

- **New data model reflects a change in philosophy to return much more metadata to the user and enable more ways to search the data sources.**



- New data model more fully describes the holdings of data providers and defines everything that a particular data provider supports (particularly requested by SunPy folks)
- New definitions that allow us to unambiguously describe data products, including derivate data products. Necessary for new data products from DKIST or PUNCH.
  - HMI 720s – Line of site magnetograms which are produces every 720s from filtergrams taken in the last 1260s.
- All parts of the new data model definitions are accessible through RESTful Registry APIs.
- Make available for community comment in November.

Data Providers



VSO Time / Instrument Search Form

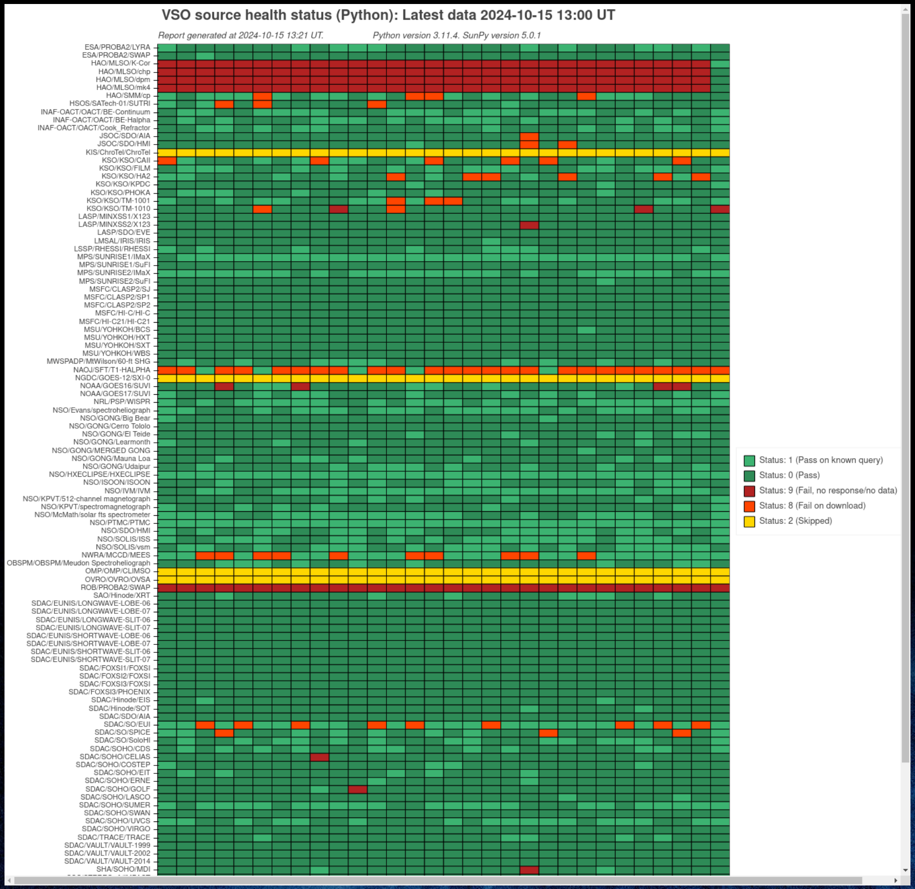
Version 1.4

All from Provider	All from	Source	Instrument	Date Range
□ ESA		PROBA2	□ LYRA	2010.03.09 →
			□ SWAP	2010.03.09 →
□ HAO	□	MLSO	□ K-Cor	2013.09.30 →
			□ chp	1996.04.20 – 2013.08.02
			□ dpm	1994.02.20 – 2010.02.23
			□ mk4	1998.10.01 – 2013.07.20
	□	SMM	□ cp	1980.03.02 – 1989.11.18
□ HSOS		SATech-01	□ SUTRI	2022.09.05 – 2022.12.05
□ INAF-OACT		OACT	□ BE-Continuum	2012.01.02 →
			□ BE-Halpha	2012.01.02 →
			□ Cook_Refractor	2009.01.02 →
□ JSOC		SDO	□ AIA	2010.05.12 →
			□ HMI	2010.03.29 →
□ KIS (unavailable)		ChroTel	□ ChroTel	2012.04.01 →
□ KSO		KSO	□ CAII	2010.07.31 →
			□ FILM	1973.05.05 – 2000.03.10
			□ HA2	2008.06.01 →
			□ KPDC	2007.07.02 →
			□ PHOKA	1989.03.27 – 2007.06.13
			□ TM-1001	1998.09.24 – 2005.07.12
			□ TM-1010	2005.07.13 – 2010.09.22
□ LASP	□	MINXSS1	□ X123	2016.04.27 – 2017.05.07
	□	MINXSS2	□ X123	2018.12.03 – 2019.01.06
	□	SDO	□ EVE	2010.04.30 →
□ LMSAL		IRIS	□ IRIS	2013.07.16 →
□ LSSP		RHESSI	□ RHESSI	2002.02.12 →
□ MPS	□	SUNRISE1	□ IMaX	2009.06.09 – 2009.06.13
			□ SuFI	2009.06.09 – 2009.06.13

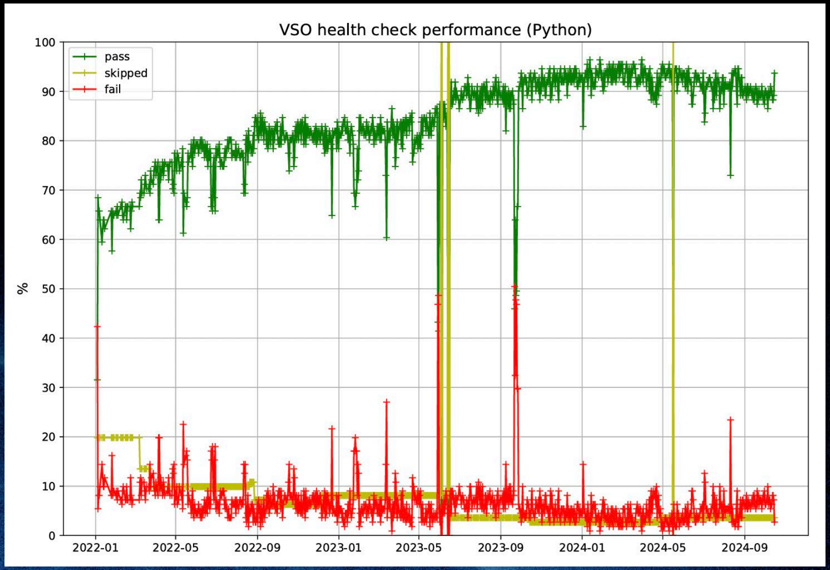
Start: 2024 Jun 22 / 22 : 00
End: 2024 Jun 23 / 01 : 59

- Recode all current Perl data providers in Python.
- Upgrade all databases to sub-second time resolution.
- Re-parse data so we can easily expose more metadata up to and including providing FITS headers.
- Expose everything that a data provider can support.
  - Ability to specify per data provider keywords outside of the core keywords
- Improved data provider framework, speed up adding new ones. (There is ~30 in the queue!)
- Maintain current data providers.

Data Provider "Health" Reports



- Currently generated once per day.
- Move to more frequent checking of data provider status.
  - Eventually move to on demand checking of data providers.



Access Protocols – *TAP* Support

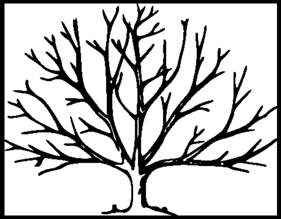
- Table Access Protocol (TAP)[1] specified by the International Virtual Observatory Alliance (IVOA)[2] defines a service protocol for accessing general table data.
- TAP+ is the ESAC Space Data Centre (ESDC)[3] extension of the Table Access Protocol.
- VSO currently uses TAP+ for searching the ESA data provider, for PROBA2 data. We are in the process of adding access to SOAR, the Solar Orbiter Archive. Expand this development to access other instruments in the ESA Heliophysics Archive (Arnaud's talk). Interfacing with VESPA (ESN-TAP) is another target.
- **In addition to searching data providers using TAP protocols, VSO 2.0 will support querying of all of the data providers it serves via TAP protocols.**

[1] https://www.ivoa.net/documents/TAP/

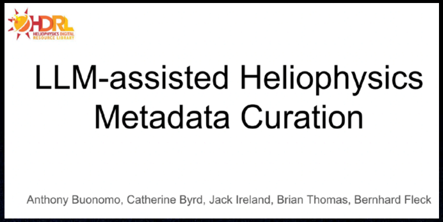
[2] https://www.ivoa.net

[3] https://www.cosmos.esa.int/web/esdc/

Other Projects



- Investigating new file formats such as Advanced Scientific Data Format (ASDF) to encapsulate all metadata (all FITS headers) in an observation set. (DKIST does this – VSO testing with AIA).
  - “I just want all the headers” – We will grant this wish!
- Integration with other projects such as “LLM-assisted Heliophysics Metadata Curation”, Data ⇔ Paper.



LLM-assisted Heliophysics Metadata Curation

Anthony Buonomo, Catherine Byrd, Jack Ireland, Brian Thomas, Bernhard Fleck

- Working with data in the cloud.

Summary

VIRTUAL SOLAR OBSERVATORY
SINCE 2002

- The VSO has been successfully serving the Solar and Heliophysics community for two decades.
- We are using that experience and the lessons we have learned, to upgrade the VSO to better serve the scientific needs of the community in the US, Europe, and around the world.
- We are also upgrading the VSO, so it continues to be sustainable, and to be maintainable.

“First, serve your community.”

VSO 2.0 Coming April-June 2025!!