Open Science Challenges in Heliophysics

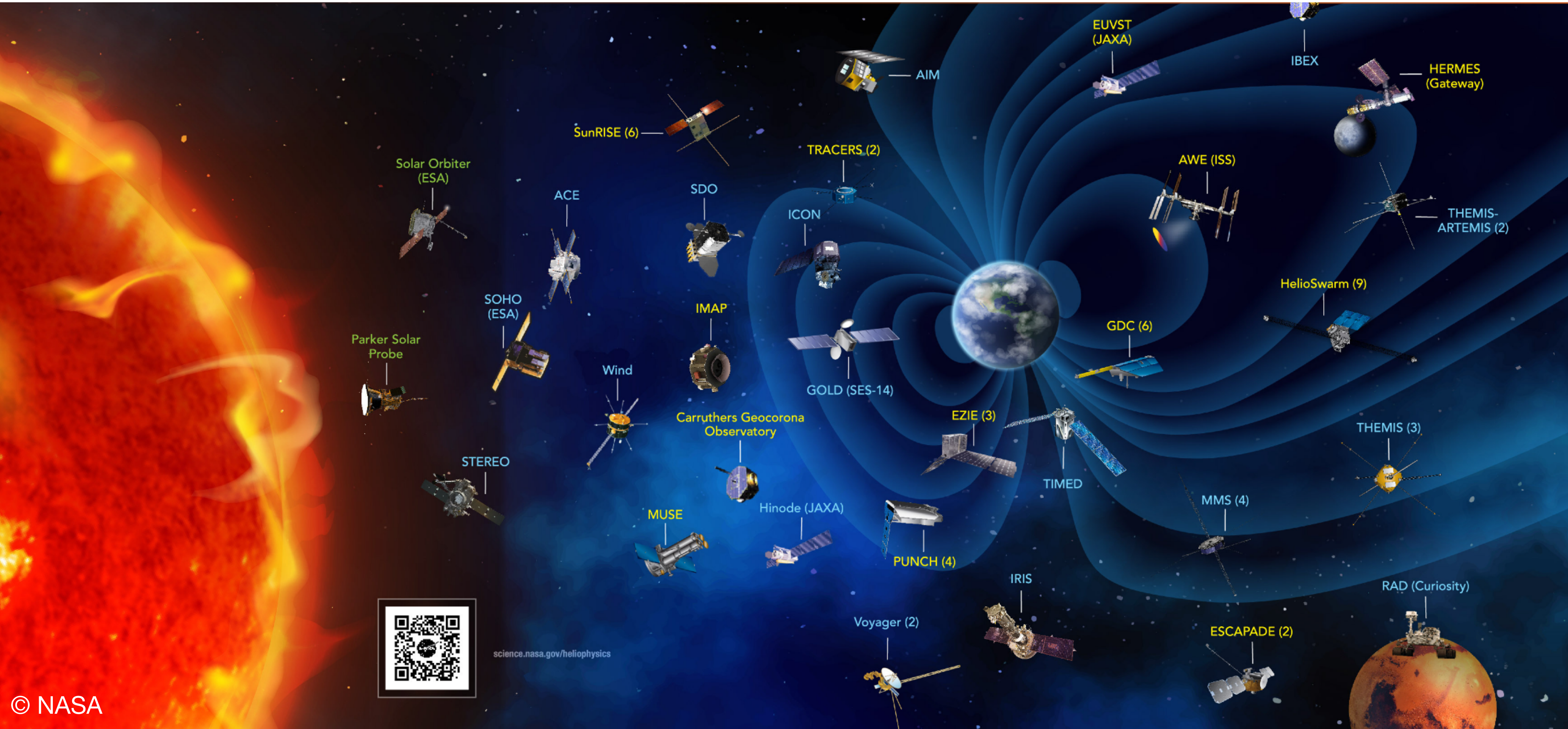
B. Cecconi

DASH — Laurel — Oct 9th 2023

Heliophysics & space physics

multipoint, multi messenger

- ground + space / single + distributed / in-situ + remote



Sharing is not an option

open science since early 1998

Studied object requires sharing data, using multiple sources.
Standardisation of accessibility is key for re-use or fundability.

- CDF+ISTP (1998): data standards for data exchange
- SPASE as a common metadata registry (2005)
- Inception of the Virtual Solar Observatory (2005)
- solar events in HEK (2007)
- FITS+WCS for solar physics (2011)
- simulation extension for SPASE (2014)
- HAPI for accessing time series (2021)

Heliophysics now

on the way to FAIR

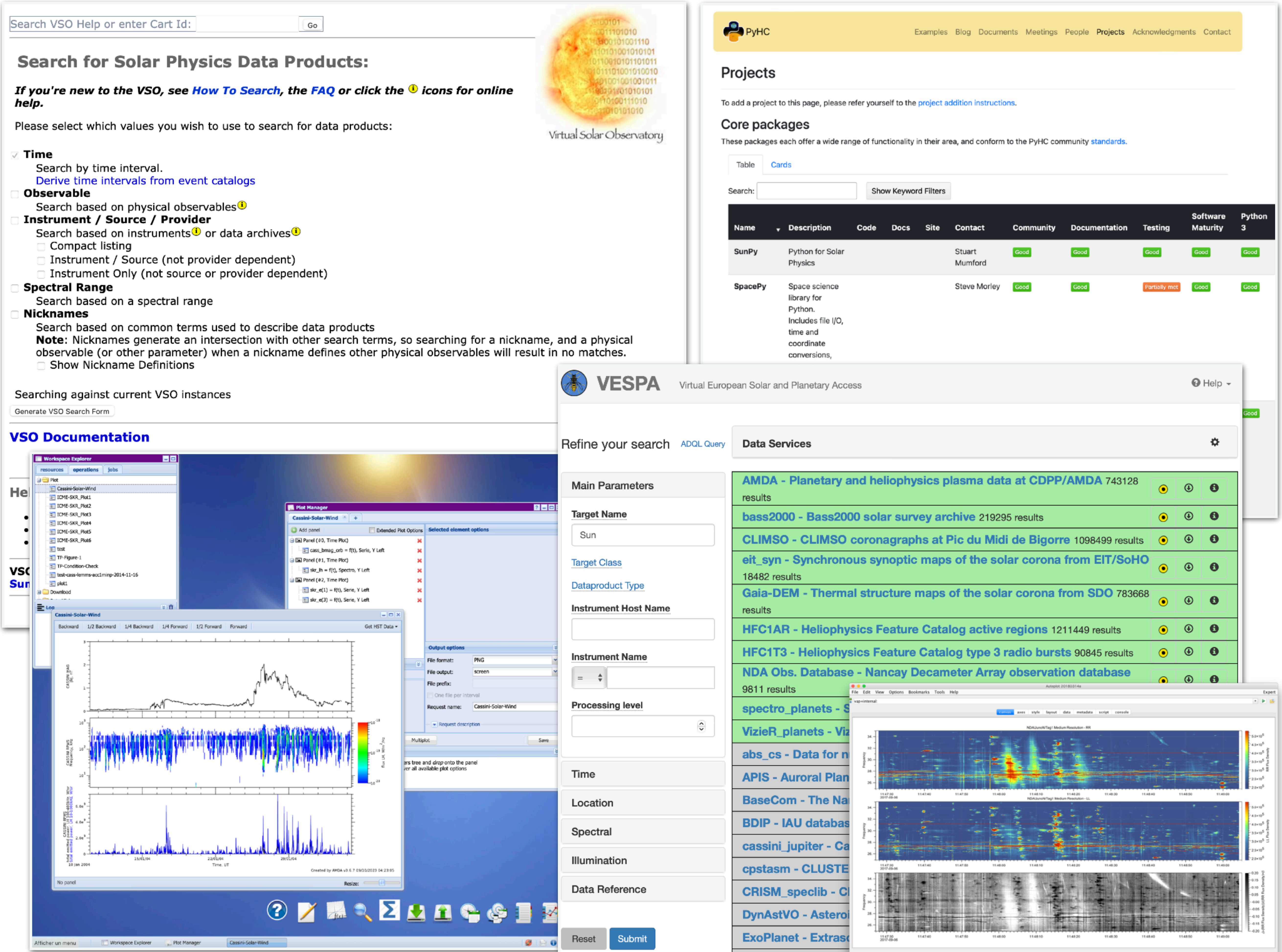
FAIR principles (Findable, Accessible, Interoperable, Re-usable)

- Findable:
  - VSO (https://sdac.virtualsolar.org/),
  - HDP (https://heliophysicsdata.gsfc.nasa.gov/)
  - VESPA (https://vespa.obspm.fr/)
  - AMDA (http://amda.cdpp.eu/)

- Accessible:
  - All data are public, with community standard format
  - Python modules to access data (PyHC)

- Interoperable
  - HAPI implements interoperable standard for time series
  - Data format (CDF, FITS)

- Reusable:
  - SPASE registry now includes DOI for data citation
  - Python modules to process data (PyHC)


Search VSO Help or enter Cart Id:
Go
Search for Solar Physics Data Products:
If you're new to the VSO, see How To Search, the FAQ or click the icons for online help.
Please select which values you wish to use to search for data products:
Time
Search by time interval.
Derive time intervals from event catalogs
Observable
Search based on physical observables
Instrument / Source / Provider
Search based on instruments or data archives
Compact listing
Instrument / Source (not provider dependent)
Instrument Only (not source or provider dependent)
Spectral Range
Search based on a spectral range
Nicknames
Search based on common terms used to describe data products
Note: Nicknames generate an intersection with other search terms, so searching for a nickname, and a physical observable (or other parameter) when a nickname defines other physical observables will result in no matches.
Show Nickname Definitions
Searching against current VSO instances
Generate VSO Search Form
VSO Documentation
Virtual Solar Observatory
PyHC
Examples
Blog
Documents
Meetings
People
Projects
Acknowledgments
Contact
Projects
To add a project to this page, please refer yourself to the project addition instructions.
Core packages
These packages each offer a wide range of functionality in their area, and conform to the PyHC community standards.
Table
Cards
Search:
Show Keyword Filters
Name
Description
Code
Docs
Site
Contact
Community
Documentation
Testing
Software Maturity
Python 3
SunPy
Python for Solar Physics
Stuart Mumford
Good
SpacePy
Space science library for Python. Includes file I/O, time and coordinate conversions,
Steve Morley
Partially met
VESPA
Virtual European Solar and Planetary Access
Help
Refine your search
ADQL Query
Data Services
Main Parameters
Target Name
Sun
Target Class
Dataproduct Type
Instrument Host Name
Instrument Name
Processing level
Time
Location
Spectral
Illumination
Data Reference
Reset
Submit
AMDA - Planetary and heliophysics plasma data at CDPP/AMDA 743128 results
bass2000 - Bass2000 solar survey archive 219295 results
CLIMSO - CLIMSO coronagraphs at Pic du Midi de Bigorre 1098499 results
eit_syn - Synchronous synoptic maps of the solar corona from EIT/SoHO 18482 results
Gaia-DEM - Thermal structure maps of the solar corona from SDO 783668 results
HFC1AR - Heliophysics Feature Catalog active regions 1211449 results
HFC1T3 - Heliophysics Feature Catalog type 3 radio bursts 90845 results
NDA Obs. Database - Nancay Decameter Array observation database 9811 results
spectro_planets
VizieR_planets
abs_cs - Data for
APIS - Auroral Plan
BaseCom - The Na
BDIP - IAU databas
cassini_jupiter - Ca
cpstasm - CLUSTE
CRISM_speclib
DynAstVO - Asteroi
ExoPlanet - Extras
Exotopo - Simulated Topography of Exoplanets 0 result
Workspace Explorer
resources
operations
jobs
Plot Manager
Cassini-Solar-Wind
Selected element options
Output options
File format:
PNG
File output:
screen
Request name:
Backward
1/2 Backward
1/4 Backward
1/4 Forward
1/2 Forward
Forward
Get HST Data
Multiplot
Save
Afficher un menu

Interplanetary use case

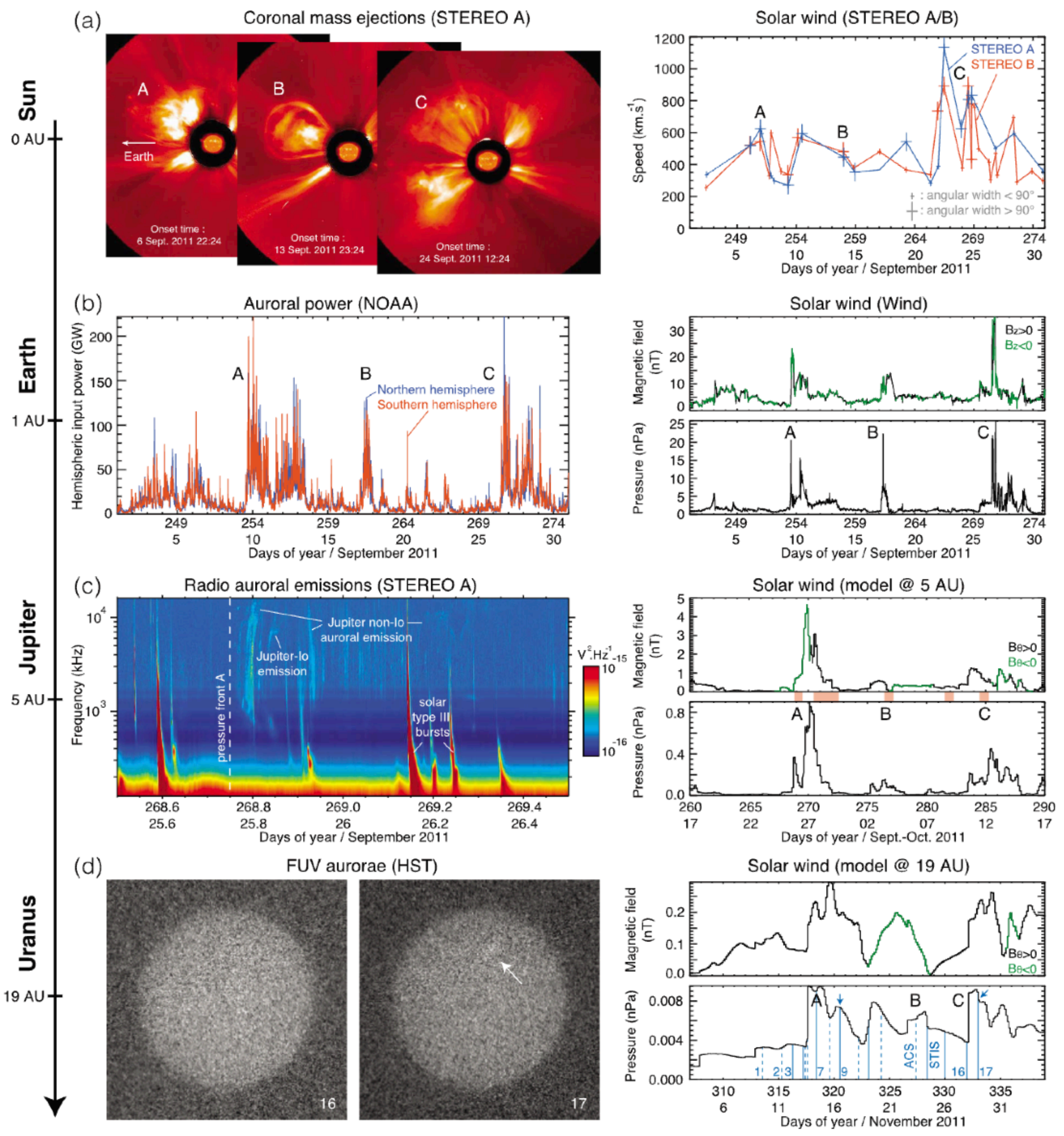
tracking event throughout the solar system

- Use case: solar event impact at planets.
  - analyze observed feature => go back to drivers
  - propagate event => test if feature is observed

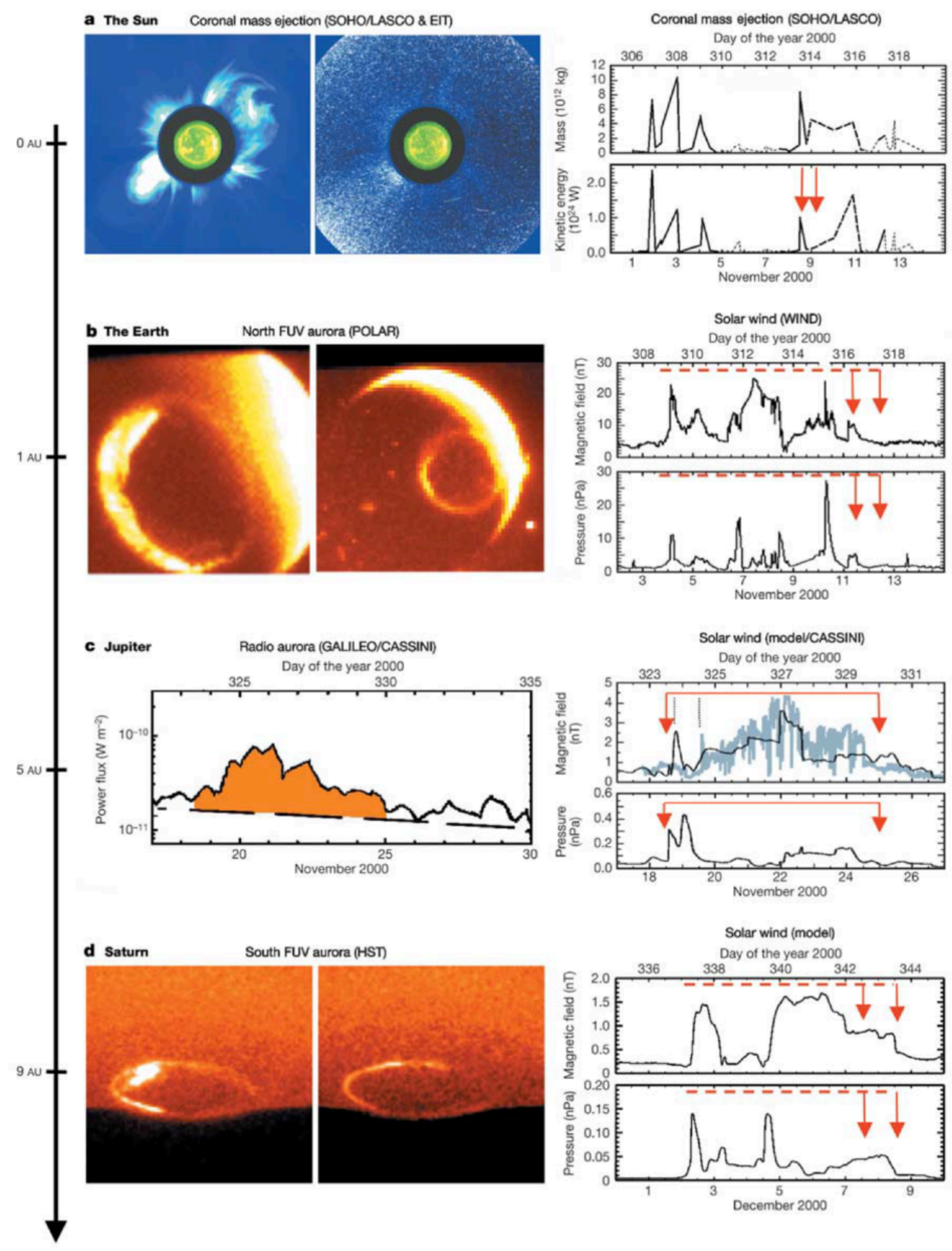
- Requires:
  - data discovery (e.g.: space in-situ date, HST data, ground radio emissions)
  - data display / processing
  - model discovery => solar wind conditions (propagation tool)
  - model input parameters => data discovery
  - running model
  - gathering results into platform with initial data

- Many studies (hand-made)
  - Cecconi et al. FrASS 9, (2022) doi:10.3389/fspas.2022.800279
  - Witasse et al. JGR 122, 7865–7890 (2017) doi:10.1002/2017ja023884
  - Lamy et al. GRL 39, (2012). doi:10.1029/2012gl051312
  - Prangé et al. Nature 432, 78–81 (2004). doi:10.1038/nature02986

- Lamy et al. (2012)



- Prangé et al. (2004)



Heliophysics data access

open data

- Heliophysics data are open access
  - images, timeseries, events…
- Tools and portals for data discovery
  - many solutions
  - need for interoperability
- Open source software (Python, IDL) for data processing
  - PyHC 🙌

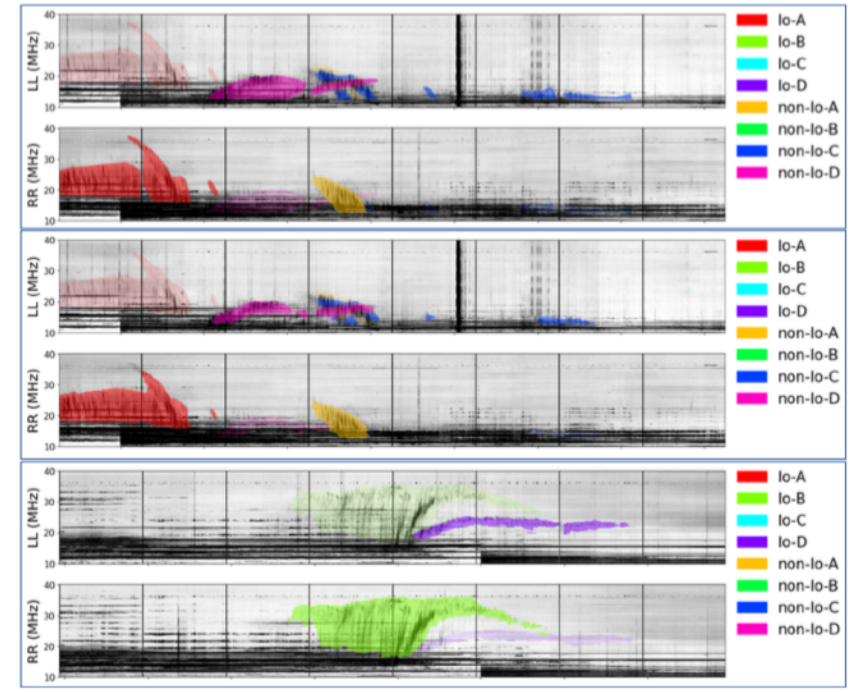
Heliophysics simulations

modelling and running

- Simulations: modelling, interpretation or prediction
- Local codes to community codes
- Opening code: for reuse (run on demand) and contribution (open source).
- Run-on-demand issues: scaling, maintenance, interfaces, storage, authorizations…
- Current status:
  - SPASE can describe models and simulation runs (IMPEx contribution)
  - CCMC or IMPEx have run-on-demand, but not standardised
  - IVOA has standardised run-on-demand interface
- Can the cloud help for scaling and authorisation ?

ML for heliophysics

event detection



- Main application of ML to heliophysics is event/feature detection. E.g.: radio burst detection (remote), solar wind event (in situ)

- Need access to
  - training set (what about rare events)
  - quality (or controlled) input data (local or stream)

- Storing results for reuse:
  - are there enough community standard for this?

- Storing provenance for reproducibility:
  - generic issue of reproducibility of ML algorithm
  - sharing code is not enough
  - sharing container with trained code?

Heliophysics tomorrow

to the cloud and beyond

- Big data requires to bring code to the data.
  - requires in turn interoperability and portability (rather than optimisation?) of code.
  - workflow management capabilities, reproducibility
  - common infrastructure for software and deployment

- NASA http://heliocloud.org/
  ESA https://datalabs.esa.int/
  many other initiatives in other domains

- Cloud of clouds:
  workflows across clouds / science platforms

IHDEA Cloud meeting (July 2023)
main highlights

- IHDEA-Cloud workshop organised in Paris (July 2023)
  with contributions from heliophysics and astronomy community

- Unconference topics:
  - *Sharing code*: metadata, testing, licensing, containers,
  - *Use Cases*: sharing notebooks, accessing platforms
  - *Discoverability*: finding, registering, shipping code
  - *Reproducibility*: containers, provenance
  - *Workflows*: workflow/pipeline manager, languages, sync/async
  - *Cross-platforms and operations*: user identity, funding, policies, orchestration
  - *User shared software stack*: seamless running of code across platforms

Summary

Heliophysics & open science

- Open science by-design since the beginning
- Core drivers: access to and reuse of data
- Standardisation of interfaces: improve or reuse
- Modelling: need to standard interfaces (description, run, access)
- ML: need quality input, community standards outputs and reproducibility challenge
- Next step: cloud with many challenges

The Europlanet-2024 Research Infrastructure project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149.