



National Aeronautics and
Space Administration



The NASA Heliophysics Digital Resource Library

IHDEA Oct 2023

Brian Thomas

POC: Brian Thomas, Project Scientist

GSFC Project Scientists: SDAC – Jack Ireland

SPDA (SPDF) – Robert Candey, Lan Jian

Data & Modeling (HDMC) – Brian Thomas

Project Manager:

HDRL – Tressa Kasulke

Oct 12, 2023



HDRL ORGANIZATIONAL CHART

HP Data and Model Consortium / HDMC

**Brian Thomas (Acting PS),
Tressa Helvey-Kasulke (PM)**

Overall management of the HDRL.

*Registries, Metadata and DOIs for all digital resources; SPASE Data Model.
Heliophysics Data Portal (HDP)*

Python and other software integration (PyHC).

Analysis and visualization services ((Py)SPEDAS, Autoplot).

Data upgrades and related services.

HelioCloud initiative with data and software from all groups.

All activities within
the various HDRL
components are
interrelated.

Space Physics Data Facility / SPDF

Robert Candey (PS), Lan Jian (DPS)

Non-solar Data Final Active Archive for NASA (and other) missions.

CDAWeb data browsing and access; Web Service access.

OMNIWeb data production and serving.

SSCWeb and 4-D spacecraft orbit facility. Common Data Format.

Solar Data Analysis Center / SDAC

Jack Ireland (PS)

*Solar Data Final Active Archive for Solar Dynamics Observatory
and other NASA missions.*

Virtual Solar Observatory data access.

Helioviewer. SolarSoft. SunPy.

High Performance Computing for NASA HP.



Collaborators

Community Coordinated Modeling Center

*Data-model comparisons; Registry of models and output; “Kamodo” enabled
visualization.*

Center for HelioAnalytics

Open Science, AI/ML Development, Mission-Enabling Tech, Community Resources, User Testing

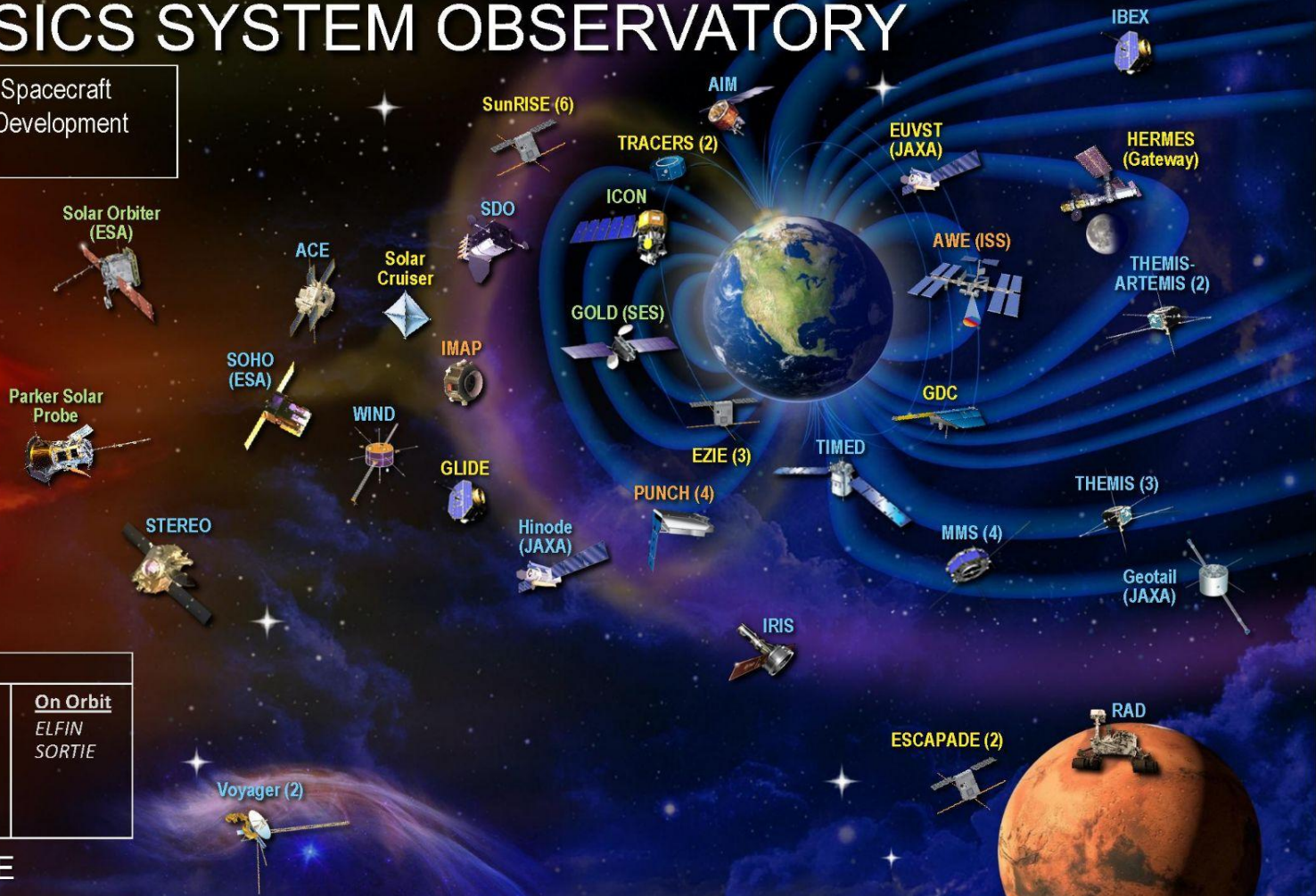
HELIOPHYSICS SYSTEM OBSERVATORY

- 20 Operating Missions with 27 Spacecraft
- 12 Missions in Formulation or Development
- 6 Under Study

■ FORMULATION
■ IMPLEMENTATION
■ PRIMARY OPS
■ EXTENDED OPS

CubeSats	
In Development	On Orbit
AEPEX	CURIE
AERO / VISTA	CuSP
CIRBE	DAILI
CODEX	Dione
CuPID	GTOSat
LAICE	LLITED
MinXSS-3	petitSat
REAL	SPORT

OPERATING & FUTURE





Vision

Where the System Observatory Comes Together

The HDRL enables the scientific analysis goals of the Heliophysics System Observatory:

- *Provisioning and curation of scientific big data* from many sources, PB volumes; (the Foundation: data, metadata, standards)
- *Support for data analysis and modeling* in multiple computational environments;
- The design and implementation of a *collaborative open science infrastructure*.

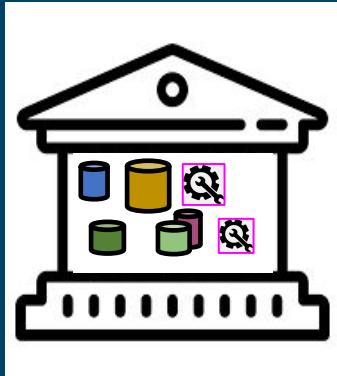
Individual missions can do great science

Unlocking groundbreaking systems science requires the HDRL

User-Driven Acceleration of Heliophysics Research: How we get there

“Preserve”

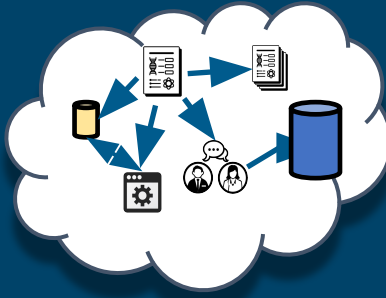
Provide Foundational Services



Maintain and upgrade existing archives and services in light of increasing demands driven by Big Data (variety & volume)

“Discover”

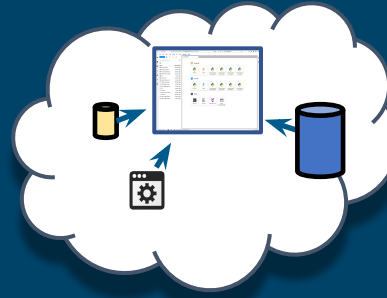
Enhance Discoverability



Increased interlinking of research artifacts, ADS integration, DOIs, improved standards, etc

“Explore Further”

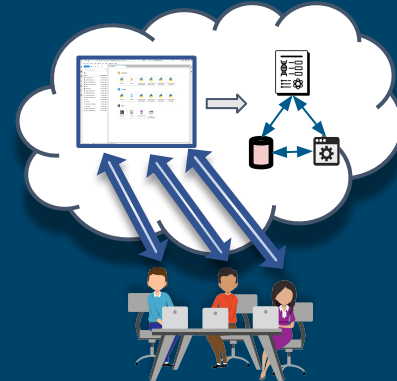
Unlock Potential



High End Compute close to large (up to ~100 TB) and Big Data (~PB) with software support (AI/ML, PyHC, etc). Leverage earth science and other platform and expertise.

“Extend and Connect”

Enable Team Open Science



Open Science; Collaborative Online Research, Compute, and Publishing Platform & Tools; Open Data; Citizen Science

Goal

Strategy

Past Year Highlights: Discover

- **Helioviewer**

- Mobile ready; CCMC flare predictions and DONKI CME/Flare notifications available; prototype model & data movie streaming

- **HelioPhysics Data Portal (HDP)**

- Added API

- **Registration improvements**

- Including SPASE (metadata) improvements, increased resourcing and reorganized to DevOps; International registry service development?

- **SMD Science Discovery Engine**

- Initiated collaboration; Heliophysics is an early enabler and available in prototype

- **ADS (SciEx) collaboration**

- Expanded collaboration to explore interlinking heliophysics data with ADS using Machine Learning

- **CDAWeb Improvement**

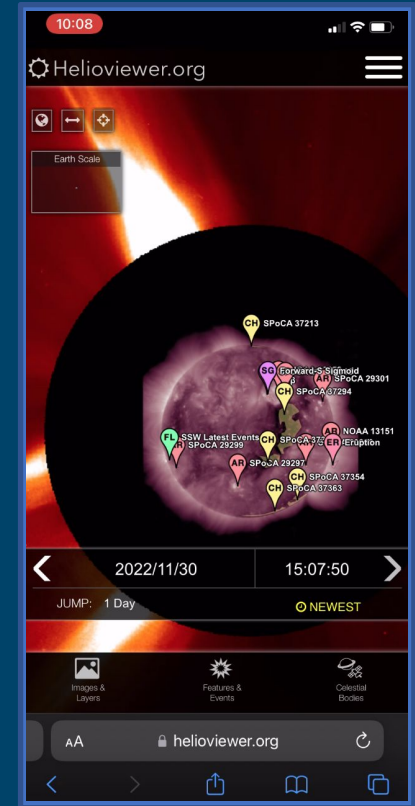
- added the link to metadata, SPASE and DOI registry for datasets, added on-the-fly data availability chart, made the axis scales adjustable.

- **4D Orbit Viewer**

- Converting Java-based orbit viewer for SSC orbits to browser-based interface

- **Accelerated HAPI adoption**

- HAPI interface to Helioviewer data



Helioviewer : screenshot of the web application running on a mobile phone.

Past Year Highlights: Explore Further

- **Helioviewer**

- Prototype 3D-aware image and model data movie web application

- **Cloud environment for Heliophysics (“HSDCloud”)**

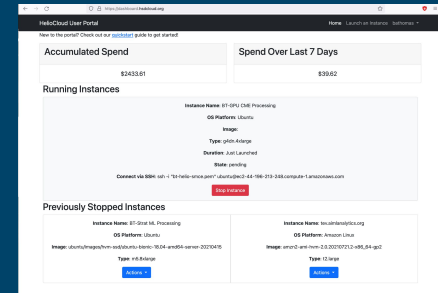
- Fully operational
- 20+ Research Science Projects (in collaboration with CfHA)
- Added GPU/ML acceleration
- Deployed “User Portal”
- Uploaded ~> 600 TB of science data to S3
- 135+ registered Users, >~ 60 active users (activity in past month)
- Supporting other efforts (Helioviewer, Helionauts, VSO, CfHA, LWSTM, SAMMER, and more!)

- **On-premises compute with NCCS**

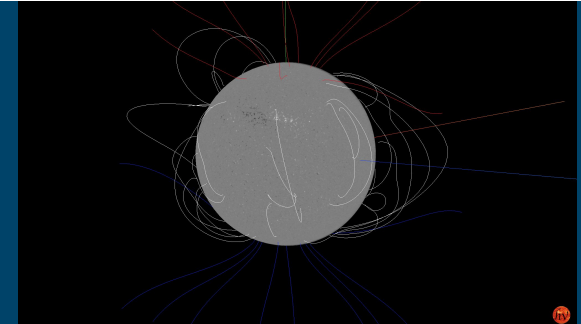
- Copying of SDAC data to NCCS
- HelioCloud data analysis environment being implemented

- **PyHC (Python in Heliophysics Community)**

- Summer School - 500 registered, HelioCloud support (~80-120)
- Now 70 projects



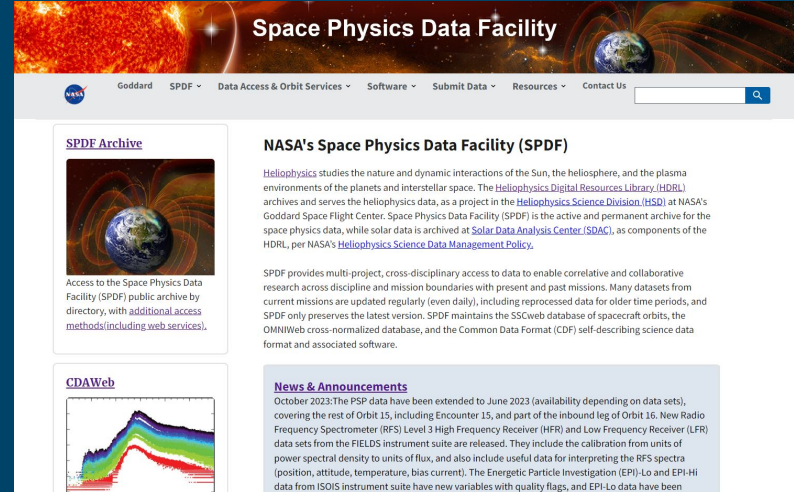
HelioCloud User Portal: Allows easy start, stop, creation, and deleting of cloud-hosted VMs for science projects and financial tracking of users resources.



New Helioviewer Capabilities: demonstration of models plus data - magnetic field model overlaid on HMI data streaming into new browser-based 3D movie client.

Past Year Highlights: Extend and Connect

- **Updating common websites**
 - New content and updated NASA look & feel
 - Adding quick-start guides for formats (e.g., CDF) and services in various software environments
- **ADS collaboration**
 - Collaboration demonstrated improved space physics searches
- **Common Python-based Heliophysics Data Analysis Environment**
 - Continued development to add GPU/ML acceleration.
 - Collaboration with PyHC, IHDEA and 2i2c communities.
- **Published contributed (ROSES-funded) data**
 - AIA active region dataset, SPASE described



Website Upgrades: Screenshot of redesigned SPDF website. Planned deployment: October 2023 for SPDF, summer 2024 for SDAC.

FY24: Strategy

- **Discovery service improvements**

Enhanced discovery site/service. Enhanced event list and API support.

- **Deploy enhanced compute capabilities**

HelioCloud, NCCS into production

- **Increase data available for analysis, discovery**

In cloud and on premise to petabyte volumes, push for SPASE evolution, CCMC/HDRL prototype

- **Engagement**

IHDEA / Workshops : to engage community / develop partnerships and shared vision / standards (SPASE, STC, ??) / collaboration (cloud, registry) / DASH 2?

- **Open Science / FAIR**

Outreach, definition of strategy, projects, workshop(s), IHDEA wg?

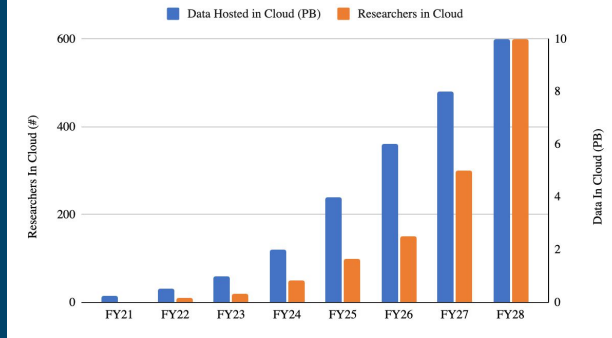
- **Conduct outreach to engage broader community**

Spread word and get broader adoption, engagement; user group

- **Deploy new infrastructure**

HelioCloud service components, other service API(s)

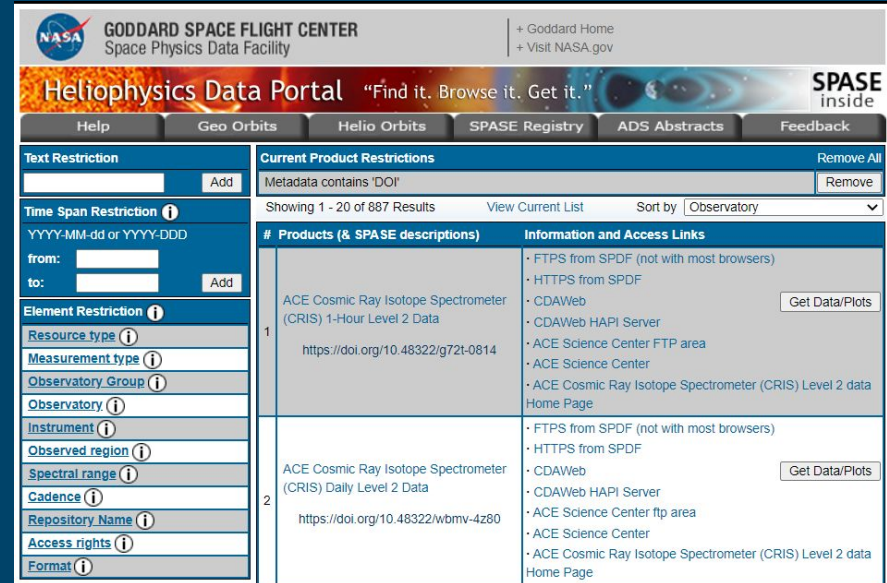
HelioCloud Capability Growth



HelioCloud: Planned growth in researchers and data hosted by the instance at NASA

FY24(+): Discover

- **HDP** — Improved user interface. Standardization of API. Search by phenomena.
- **HEK** — Ingestion / full support for non-solar event data.
- **ADS** — Integration of software, other research artifacts to interlink with NASA datasets and Heliophysics publications
- **Improved support for Event Lists**
- **HelioCloud** - develop science database, data registry searches across instances.
- **CDAWeb data browser** - add interactive search and display.



The screenshot shows the NASA Goddard Space Flight Center Space Physics Data Facility Heliophysics Data Portal. The interface includes a search bar with "Find it. Browse it. Get it." and a "SPASE inside" logo. Navigation tabs include Help, Geo Orbits, Helio Orbits, SPASE Registry, ADS Abstracts, and Feedback. A "Text Restriction" field is present with an "Add" button. Below it are "Time Span Restriction" (YYYY-MM-dd or YYYY-DDD) and "Element Restriction" (Resource type, Measurement type, Observatory Group, Observatory, Instrument, Observed region, Spectral range, Cadence, Repository Name, Access rights, Format) sections, each with an "Add" button. The "Current Product Restrictions" section shows "Metadata contains 'DOI'" with a "Remove" button. The main content area displays search results for "Showing 1 - 20 of 887 Results", sorted by "Observatory". Two product entries are visible:

#	Products (& SPASE descriptions)	Information and Access Links
1	ACE Cosmic Ray Isotope Spectrometer (CRIS) 1-Hour Level 2 Data https://doi.org/10.48322/g72t-0814	<ul style="list-style-type: none"> • FTPS from SPDF (not with most browsers) • HTTPS from SPDF • CDAWeb • CDAWeb HAPI Server • ACE Science Center FTP area • ACE Science Center • ACE Cosmic Ray Isotope Spectrometer (CRIS) Level 2 data Home Page Get Data/Plots
2	ACE Cosmic Ray Isotope Spectrometer (CRIS) Daily Level 2 Data https://doi.org/10.48322/wbmv-4z60	<ul style="list-style-type: none"> • FTPS from SPDF (not with most browsers) • HTTPS from SPDF • CDAWeb • CDAWeb HAPI Server • ACE Science Center ftp area • ACE Science Center • ACE Cosmic Ray Isotope Spectrometer (CRIS) Level 2 data Home Page Get Data/Plots



FY24(+): Preserve

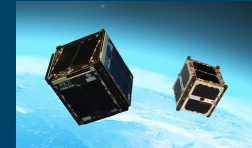
- **Enhanced data ingest and registration**

- Develop more *streamlined registration* (DOIs)
- Develop *improvements to metadata standards* to support requirements
 - Better support registration of software
 - Better support for open science publishers
- Support data from *smallsats, balloons, rockets, ground-based observatories.*

- **Transfer of Multi-PetaByte Data to SDAC**

- SDO Level 0 data (~6PB) from SDO JSOC to SDAC.

- **Offline backup of all data currently at SDAC**



FY24(+): Explore Further

• HelioCloud

- Release easily deployed cloud-based deployable environment (Blue box), AGU 2023
- Ingest up to 8 PB of data to cloud

• On premise research environment for Big Data

- In production at NCCS (Green box)
- Implementation of DaskHub environment

• Transfer of AIA Level 1 data from SDO JSOC to NCCS

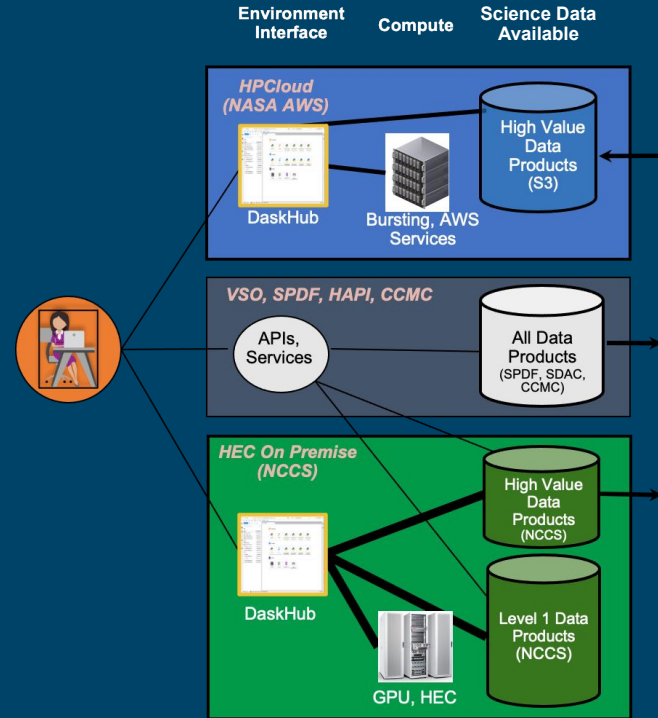
- Scientific use of AIA Level 1 data on NCCS computational environment using HelioCloud shared data analysis environment and DaskHub platform (deployed at NCCS)

• Helioviewer

- Add capability to show selected content from CCMC and SPDF

• Support AI/ML-ready data, Model / Data comparison

- With priority for cloud-based analysis by public, CCMC collaboration



New Compute Environments: In operation in the Cloud (Blue) and @ NCCS (Green) with common software and platform (DaskHub). Thickness of lines indicate throughput.

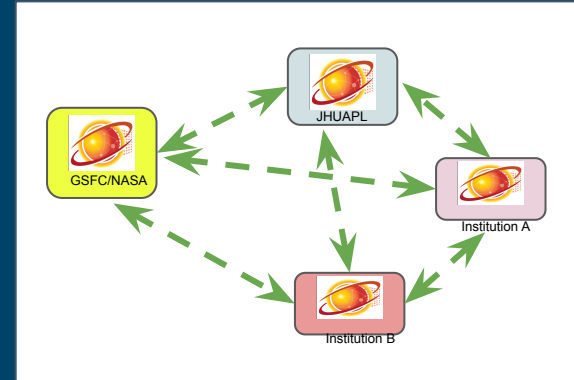
FY24(+): Extend and Connect

• HelioCloud

- Provide easily deployed framework for deployment at other institutions
- Add tools to add researchers to publish Open Science
- Add support for sharing notebooks, containerized software and other research artifacts
- Develop citizen science projects hosted in HelioCloud

• Open Science Software Publication

- Shared software environment for Heliophysics research (Python)
- Community engagement to develop best practices, standards for software publication



***HelioCloud:** Research artifacts may be easily 'published' and shared with others at other institutions.*





End