WEAVE Data Orchestration

The journey from target selection to scientific discovery

Alireza Molaeinezhad, Mike Irwin, Giuseppe D'Ago, David Murphy, Anke Ardern-Arentsen, Thomas Hajnik and Nicholas Walton

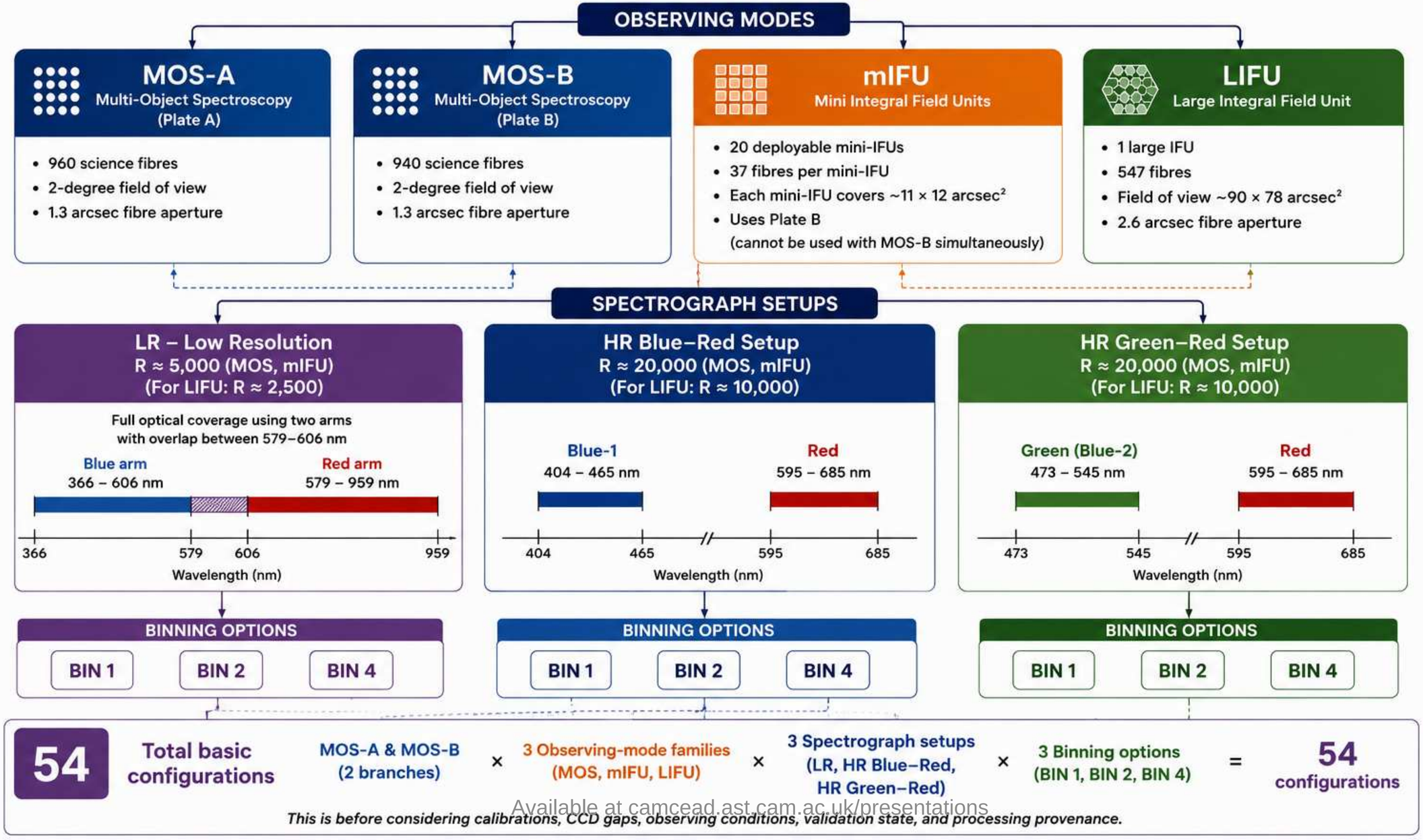
Cambridge Centre of Excellence for Astronomical Data (CAMCEAD)
Institute of Astronomy, University of Cambridge
camcead.ast.cam.ac.uk

EAS2026, Lausanne, July 2026

Available at camcead.ast.cam.ac.uk/presentations

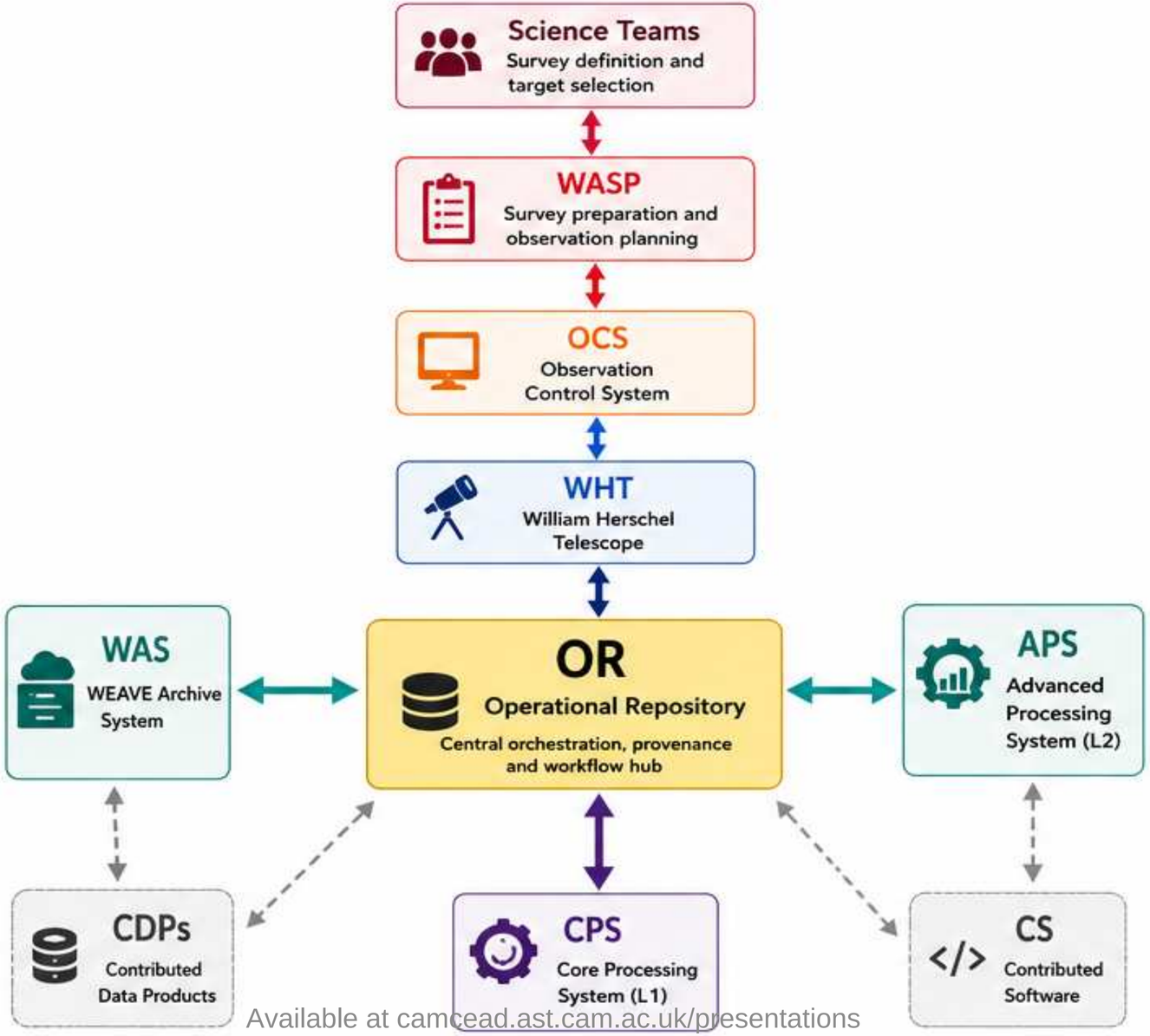
What Makes WEAVE a complex datadflow?

Multiple observing modes, spectrograph setups and binning options create a large configuration space



WEAVE Dataflow Architecture

A distributed ecosystem for survey preparation, observation execution, processing, validation and archive delivery.



WASP

WEAVE Automatic Submission Platform

From survey design to executable observations

WASP *validates, organises* and *transforms* survey requirements into executable WEAVE observations.

WEAVE OR & Data Access Point

The backbone of the WEAVE dataflow

WEAVE Operational Repository (WEAVE OR)

Central orchestration hub of the WEAVE ecosystem

- **Coordinates** data exchange across all WEAVE subsystems
- **Tracks** processing, validation and archive states in real time
- **Maintains** provenance, workflow history and data lineage
- **Triggers** automated actions and notifies data availability
- **Ensures** consistency, traceability and operational efficiency

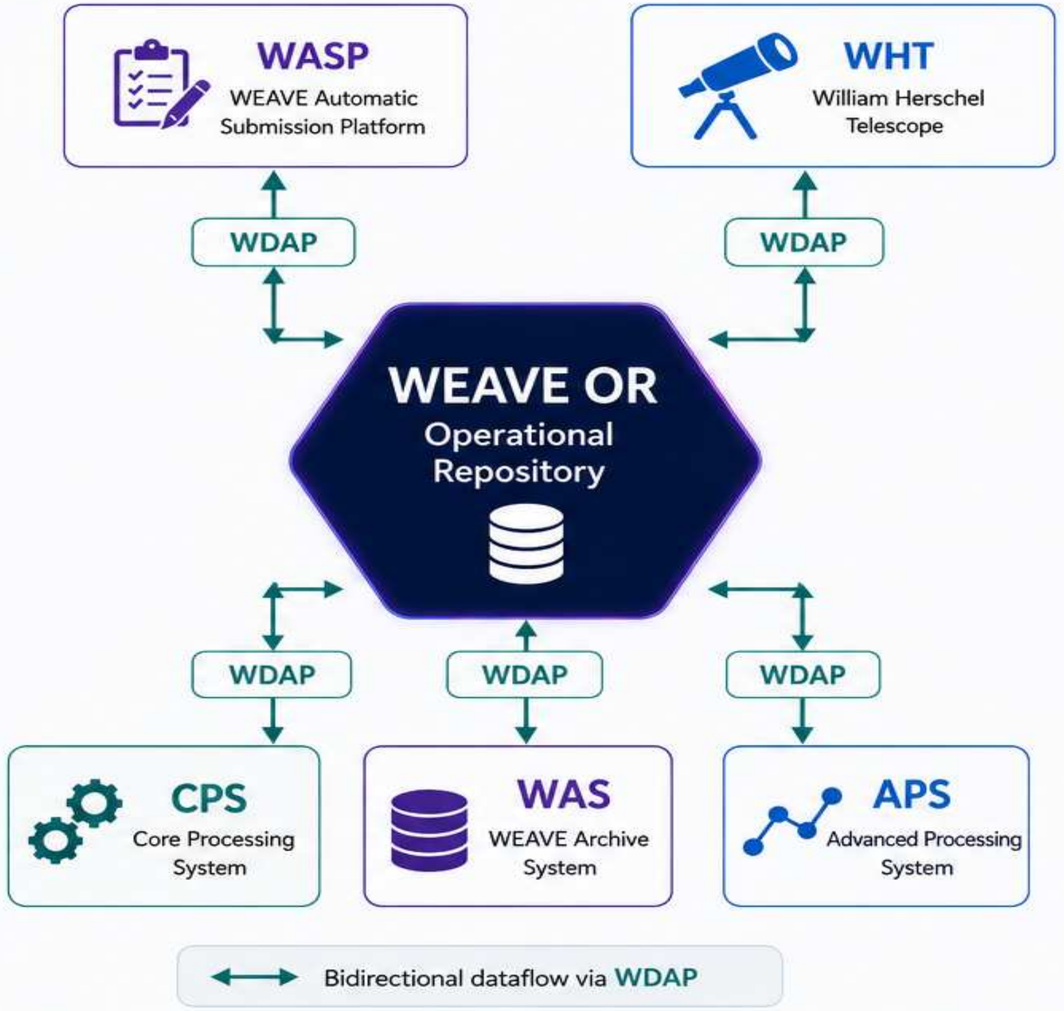
WEAVE Data Access Point (WDAP)

Recipe-driven interface between pipelines and WEAVE OR

- **Standard** interface for pipelines to interact with WEAVE OR
- Encapsulates workflow logic into reusable recipes
- **Enables** automated, scalable and reliable processing
- Provides a common framework for CPS, APS and WAS

Together, **WEAVE OR** and **WDAP** ensure a **coordinated**, **traceable** and **efficient** dataflow across the entire WEAVE system.

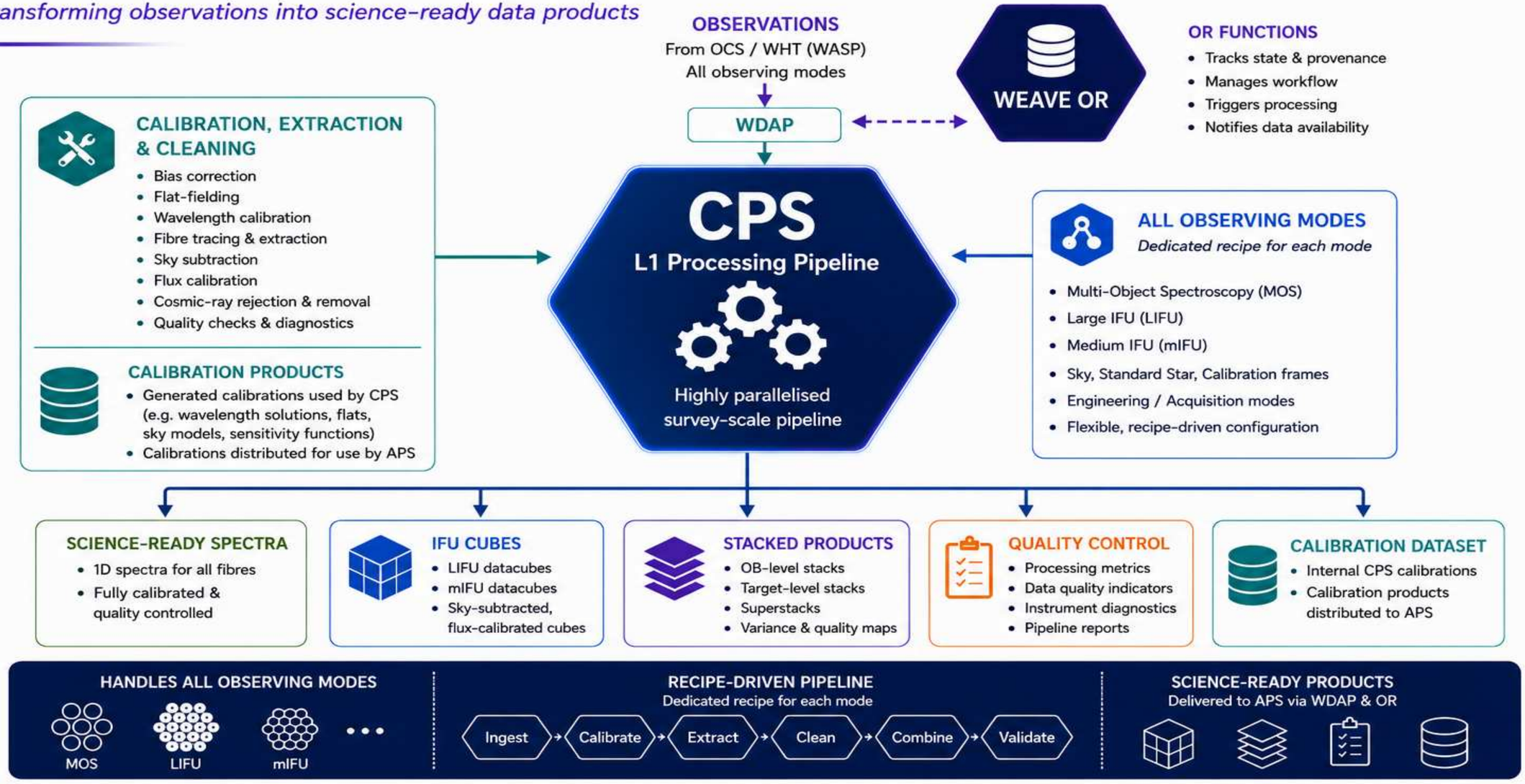
WEAVE DATAFLOW ARCHITECTURE



WEAVE OR orchestrates the dataflow and maintains the operational state.
WDAP provides the standard, **recipe-driven** interface connecting all WEAVE systems.

CPS — Core Processing System (L1)

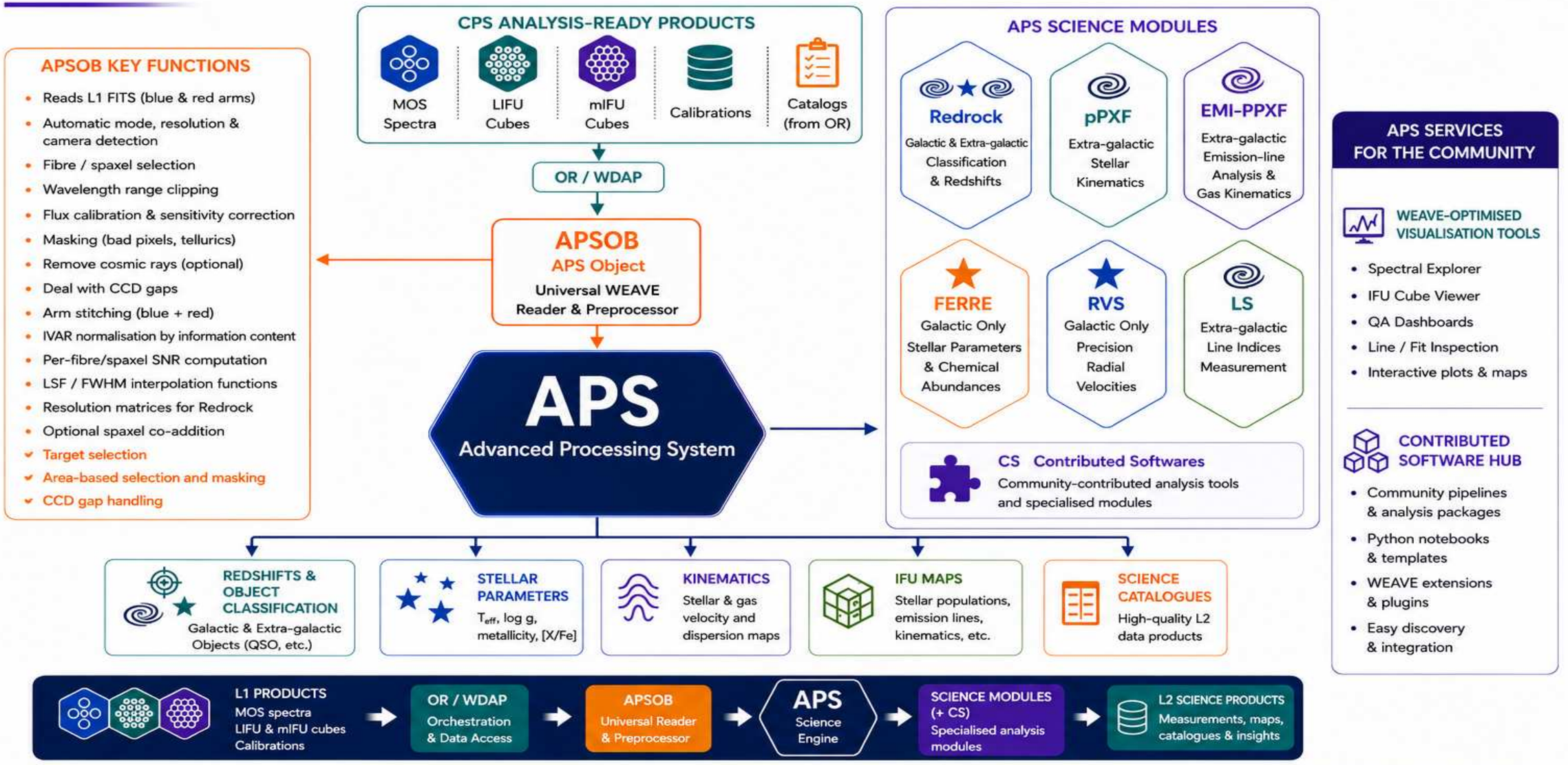
Transforming observations into science-ready data products



CPS is the high-throughput Level-1 pipeline that handles **all WEAVE observing modes** and produces **science-ready, calibrated** data products, cubes, stacks and **calibrations** — ready for scientific analysis by **APS**.

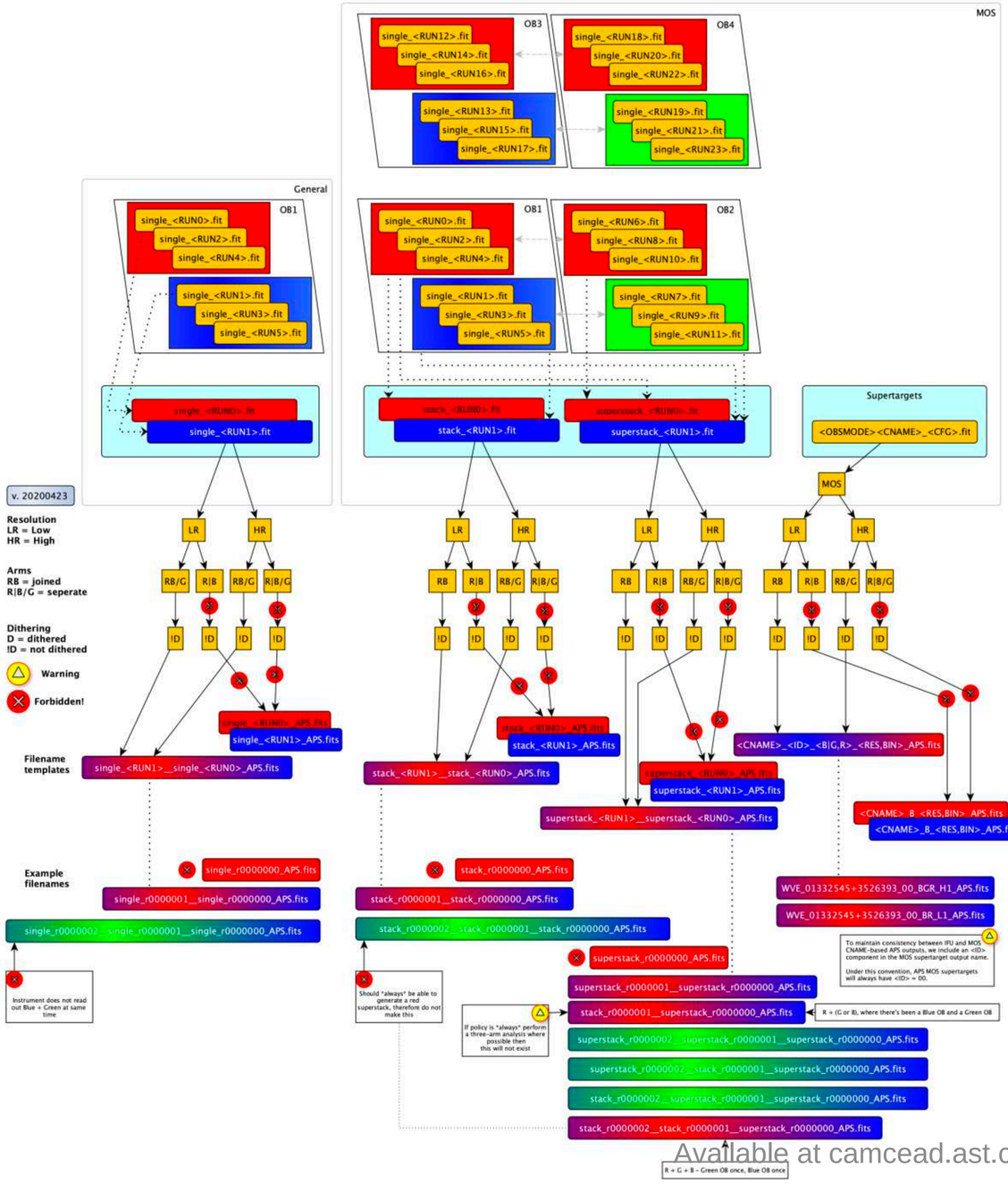
APS — Advanced Processing System (L2)

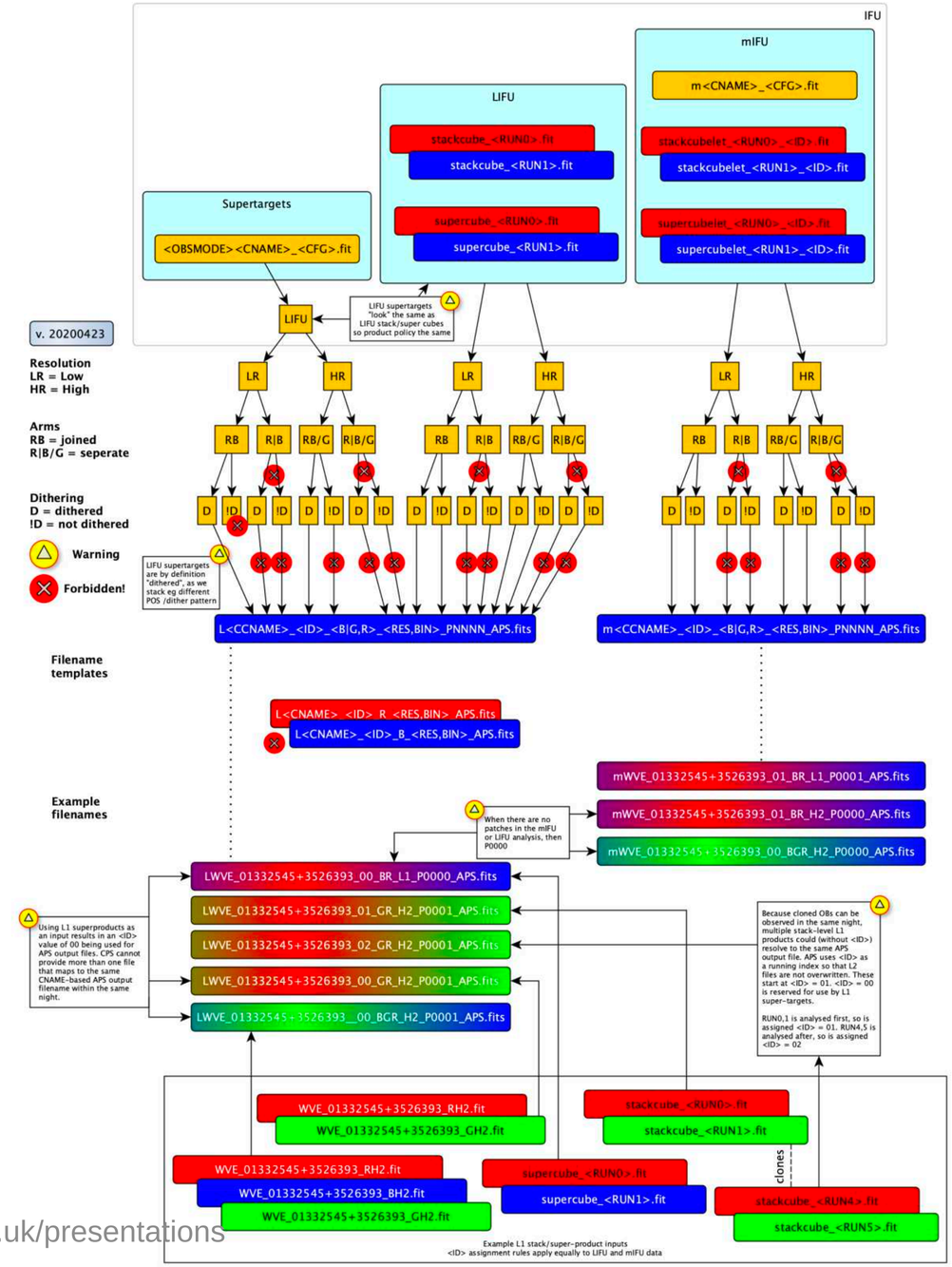
Transforming WEAVE observations into science-ready measurements, maps and catalogues



APS transforms calibrated WEAVE spectra and cubes into **science-ready** measurements, maps and catalogues.
APSOB provides a common analysis foundation for all APS modules, while APS delivers **visualisation tools** and hosts **community-contributed software.**

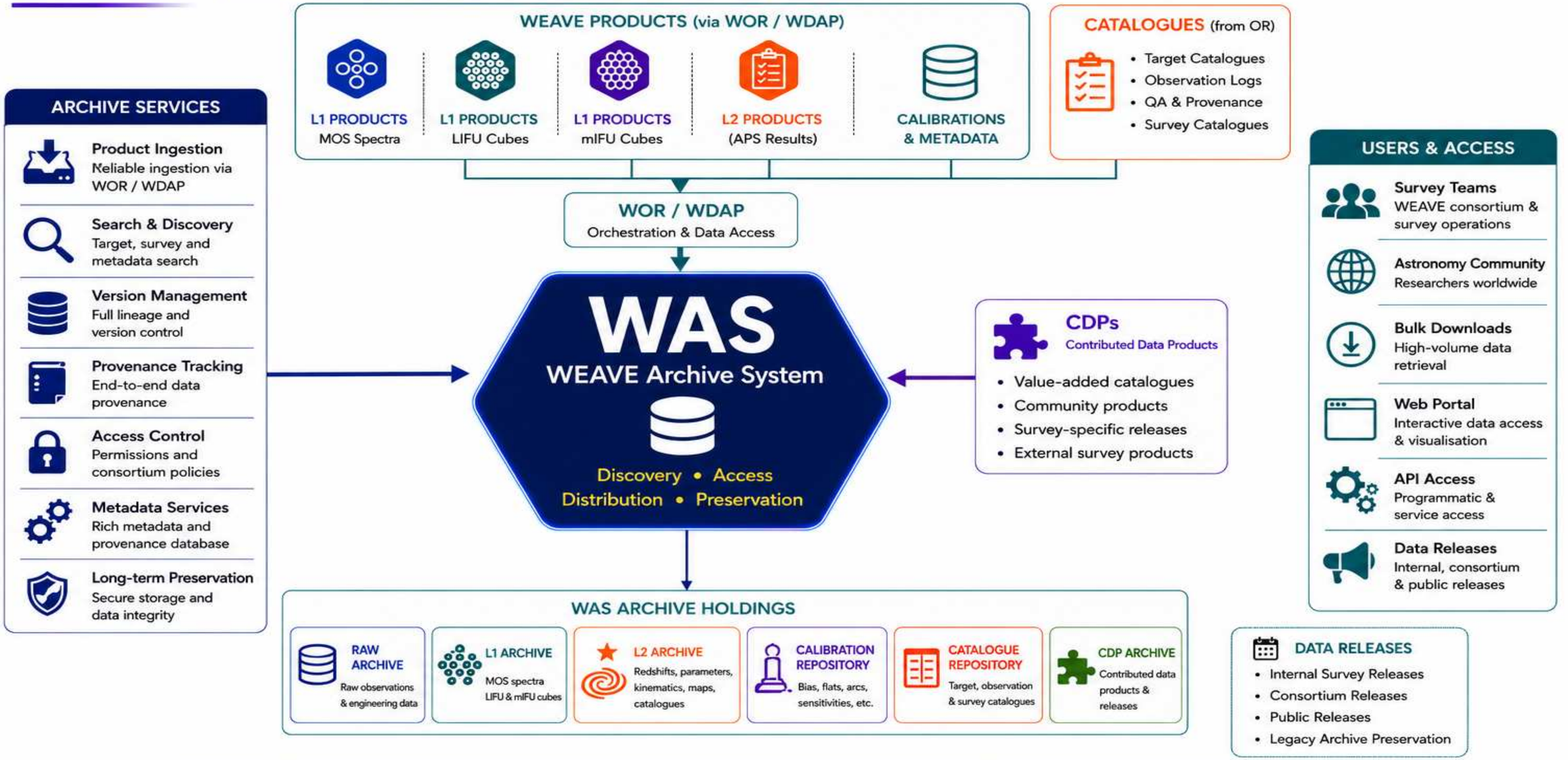
CPS/APS data IO pathways





WAS — WEAVE Archive System

Long-term preservation, discovery and distribution of WEAVE data products



WAS is the long-term archive and distribution platform for WEAVE, providing discovery, access and preservation of L1, L2, calibration, catalogue and community-contributed data products.

Daily WEAVE Data Lifecycle

From observations at the WHT to L1/L2 product availability in the WEAVE Archive System

Typical daily cycle
~7–8 hours end-to-end

La Palma, Canary Islands	Cambridge, UK	Cambridge, UK	CASU–IRIS Cloud Infrastructure	Cambridge, UK	Canary Islands, Spain
WHT Observatory					

1 Observations Completed

Overnight observations completed at the WHT (La Palma, Canary Islands)

Raw data ready for transfer

2 Data Received at CAMCEAD

Data transferred from WHT to CAMCEAD (Cambridge)

Data received at ~12:00 UK time

3 CPS Processing & Calibration

CPS processes incoming data and generates necessary calibrations

(~1 hour)

WDAP

4 Transfer via WDAP to WEAVE OR

WDAP

Data transfer via WDAP protocol to WEAVE OR

5 L1 Validation against WEAVE Data Model

L1 products validated against the WEAVE data model

(~30 min)

WDAP

6 Transfer via WDAP to APS

APS

Data transfer via WDAP protocol to APS

7 APS L2 Processing

APS processes data and generates L2 products

(≤6 hours)

WDAP

8 WDAP Transfer to WOR for Validation & Ingestion

L2 products transferred via WDAP to WOR for validation and ingestion

9 L1 & L2 Products Available in WAS

WEAVE OR signals WAS (Canary Islands) about availability of both L1 and L2 products

Previous Night	12:00	~1 hour	13:00	~30 min	13:30	≤6 hours	19:30	20:00	20:00+
Observations at WHT (La Palma)	Noon UK Time Data received at CAMCEAD		CPS processing completed		L1 validation completed		APS processing completed	WOR validation & ingestion completed	L1 & L2 products available in WAS (Canary Islands)

Daily Cycle Summary

- Data received at CAMCEAD (Cambridge) at ~12:00 UK time
- CPS processes and calibrates incoming data (~1 hour)
- Data transferred via WDAP to WEAVE OR
- L1 validated against WEAVE data model (~30 min)
- Data transferred via WDAP to APS
- APS processes and generates L2 products (≤6 hours)
- L2 products transferred via WDAP to WOR for validation and ingestion
- WEAVE OR notifies WAS (Canary Islands) of L1 & L2 availability

- La Palma, Canary Islands (WHT)
- Cambridge, UK (CAMCEAD, CPS, OR, WOR)
- CASU–IRIS Cloud Infrastructure (APS)
- Canary Islands, Spain (WAS)

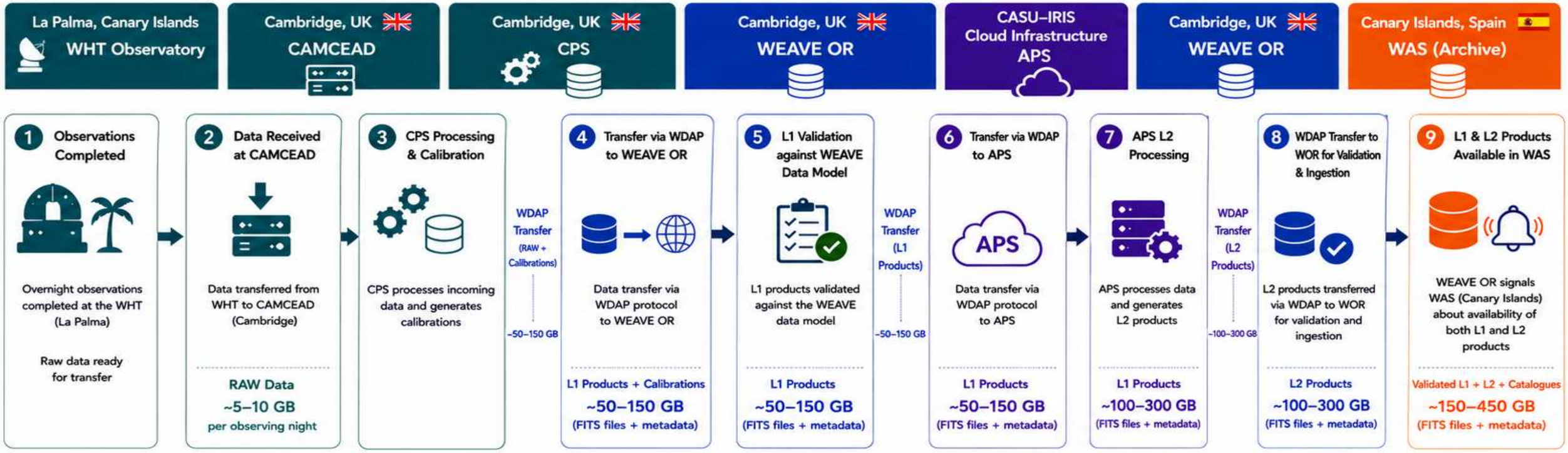
Supertargets (Monthly/Semester)

- Target-level stacking performed by CPS
- Multiple observations combined into deep Supertarget products
- Built on a monthly / semester basis
- Independent of the daily processing cycle

Typical Data Volumes Across the WEAVE Dataflow

Daily data products transferred between WEAVE processing systems

Volumes are typical daily averages and depend on observing mode (MOS, LIFU, mIFU), spectral resolution and survey configuration.

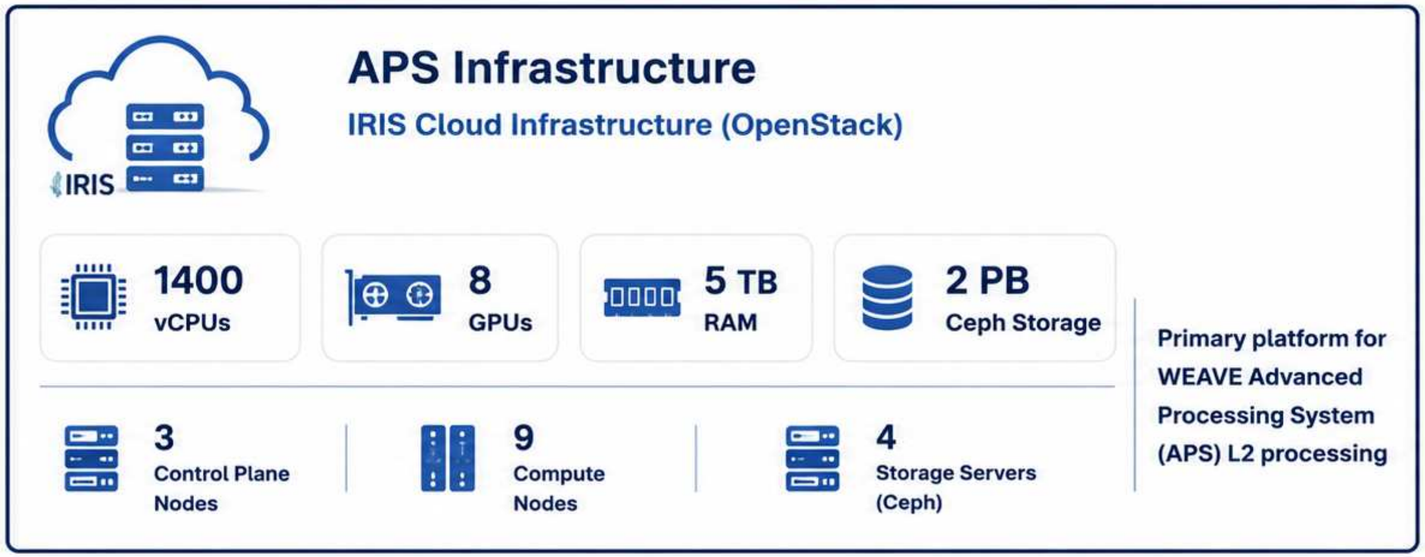


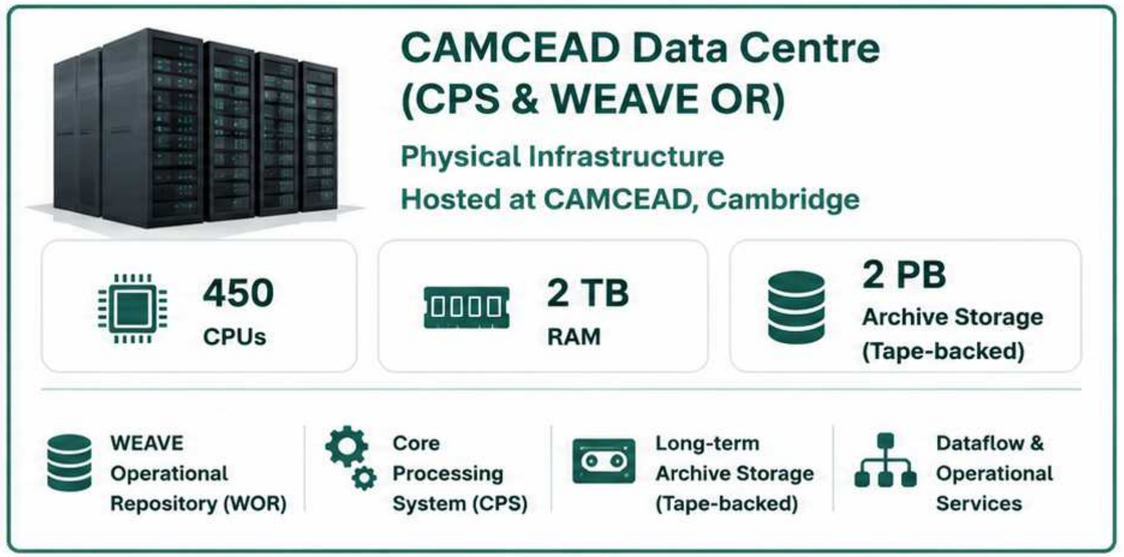
Data Volumes Between Nodes

WHT → CAMCEAD (RAW Data)	CPS → WEAVE OR (via WDAP)	WEAVE OR → APS (via WDAP)	APS → WEAVE OR (via WDAP)	WEAVE OR → WAS (via WDAP / Internal)
~5–10 GB per observing night	L1 Products + Calibrations ~50–150 GB (FITS files + metadata)	L1 Products ~50–150 GB (FITS files + metadata)	L2 Products ~100–300 GB (FITS files + metadata)	Validated L1 + L2 + Catalogues ~150–450 GB (FITS files + metadata)

All data products are transferred as FITS files and associated metadata (tables, QC, logs, provenance). Volumes depend on observing mode (MOS, LIFU, mIFU), spectral resolution and survey configuration. Values shown are typical daily averages.

WEAVE dataflow infrastructure



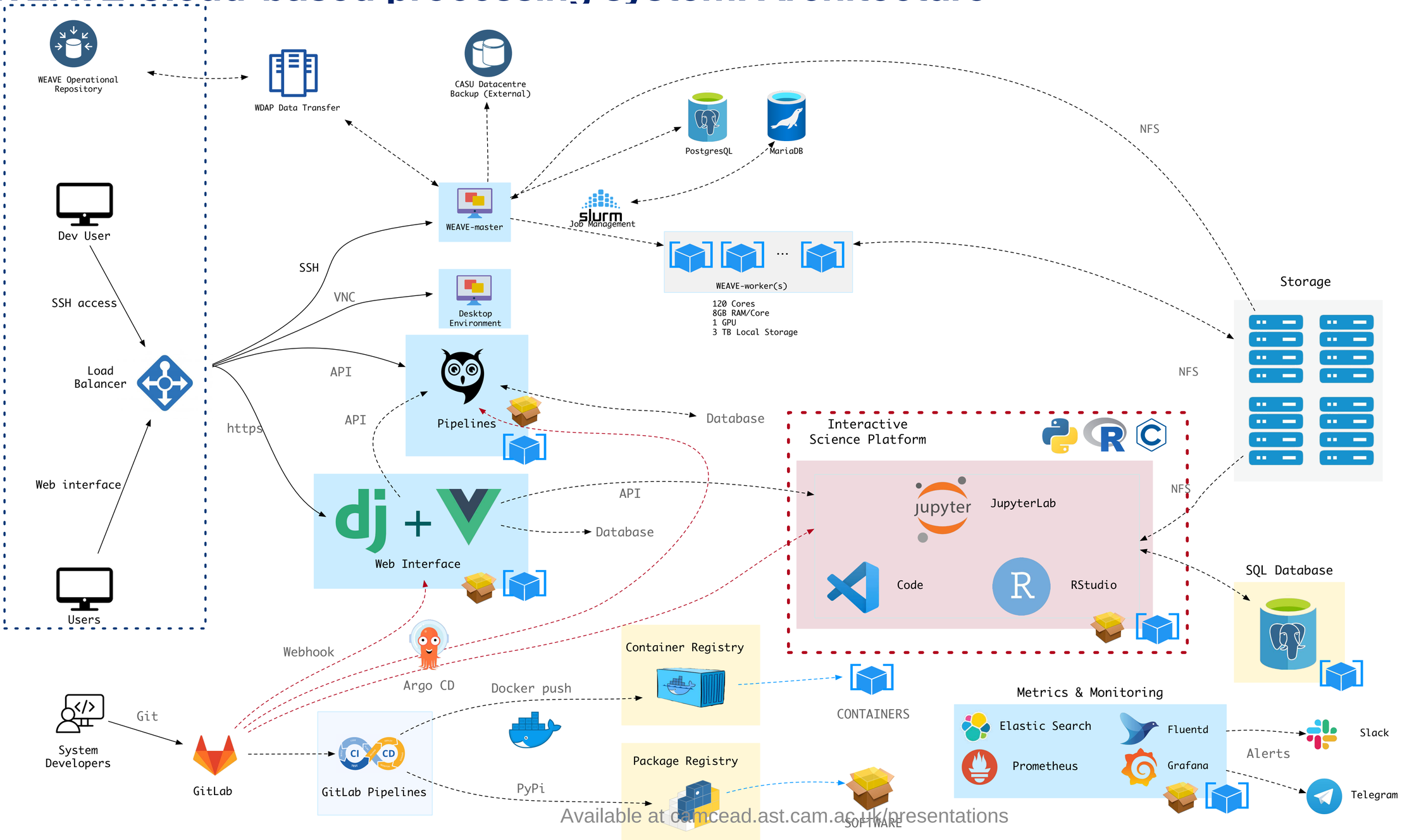


Cost Comparison: CAMCEAD vs Commercial Providers

Facility	Mode	RAW/night (GB)	L1/night (GB)	L2/night (GB)	AWS Compute cost/ day *	AWS Storage cost/day *
WEAVE	MOS-LR	1.65	8.25–16.5	33	€1500	€1600
WEAVE	MOS-HR	3.0	15–30	60		
WEAVE	LIFU LR	4.5	22.5–45	100–130		
WEAVE	LIFU HR	18.0	90–180	300–550		

*Cost estimates based on March 2026 pricing

WEAVE Cloud-based processing system: Architecture



CAMCEAD (Cloud-based) Analysis Platform and Compute access model

Direct SSH-accessible instances with SLURM:

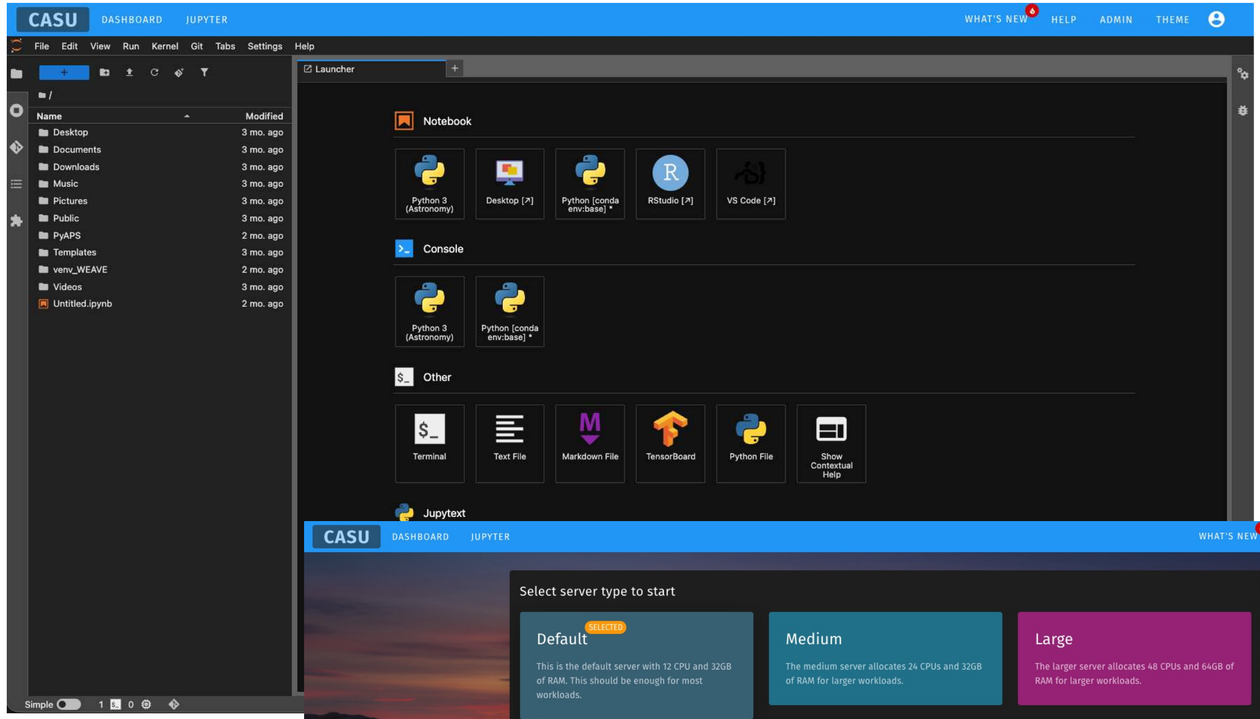
- Traditional HPC-style SSH access for batch processing.
- SLURM workload manager for job scheduling and resource allocation.
- Automated pipeline execution and large-scale processing tasks.

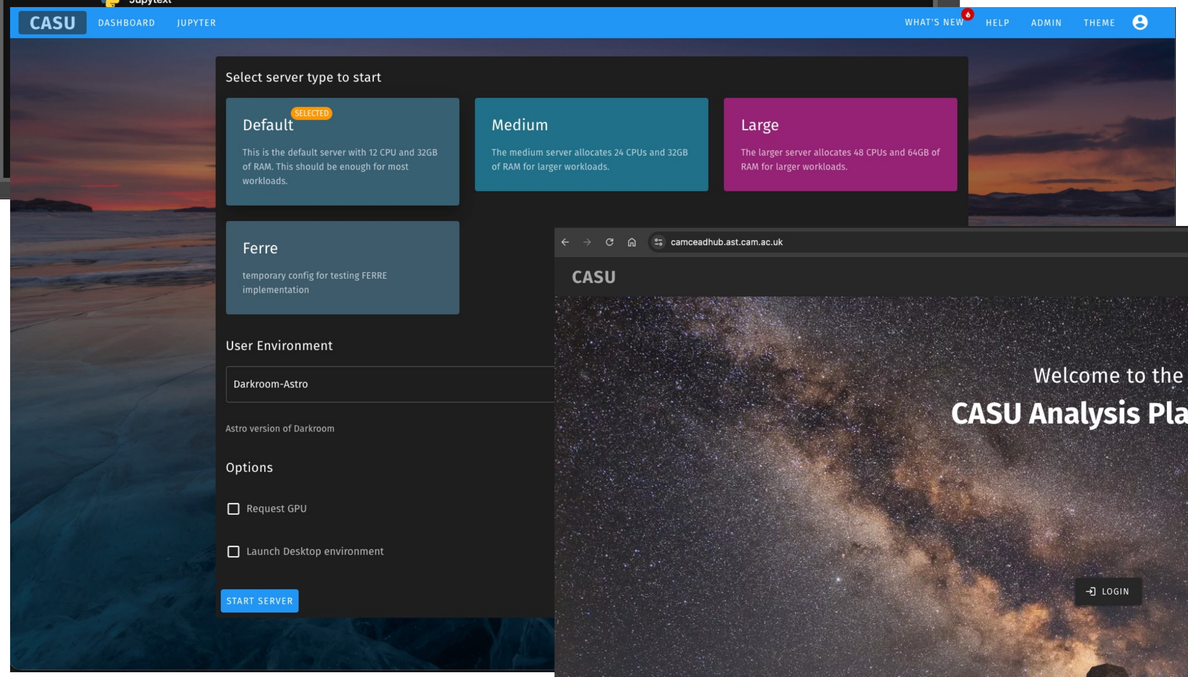
JupyterHub platform with web interface:

- Browser-based JupyterLab with configurable CPU, RAM, GPU resources.
- Pre-installed scientific stack (AstroPy, NumPy, TensorFlow, PyTorch).
- Supports up to 30 concurrent users with dynamic allocation.

Remote desktop environment (noVNC):

- Web-based graphical desktop, with all essentials for astronomy visualization tools (DS9, TOPCAT, Napari)
- GPU-accelerated rendering for large datasets





Available at camcead.ast.cam.ac.uk/presentations

Key Takeaways

- WEAVE operates as a **fully integrated end-to-end ecosystem**, connecting observation planning, data acquisition, processing, validation and archival services.
- A combination of **WOR** and **WDAP** provides centralised workflow orchestration, automated validation and secure communication between all processing systems.
- **CPS** and **APS** transform raw observations into calibrated data products, advanced scientific measurements and survey-ready catalogues within hours of data acquisition.
- The platform is designed to support the full diversity of WEAVE observations, from large MOS surveys to spatially resolved IFU observations.
- A distributed infrastructure combining the **IRIS cloud platform**, **CAMCEAD operational services** and the **WEAVE Archive System** delivers the scalability required for long-term survey operations.
- The ecosystem enables rapid data availability, reproducible processing, complete provenance tracking and sustainable access to WEAVE data products for the wider astronomical community.

Related Poster

📌 **The Performance of WEAVE's Stellar Analysis Modules**

Dr. Anke Ardern-Arentsen [Link]

For further insight into the scientific analysis and performance of the WEAVE stellar processing pipeline, please visit the accompanying poster.