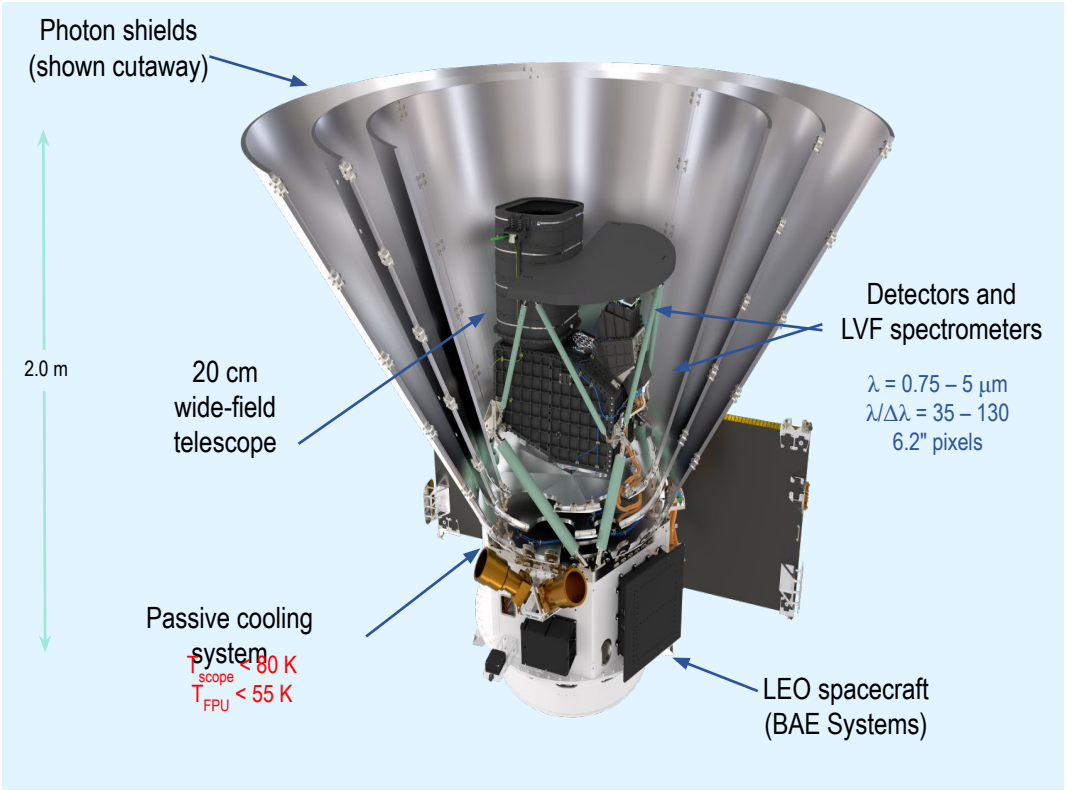
# SPHEREx and the VO at IRSA

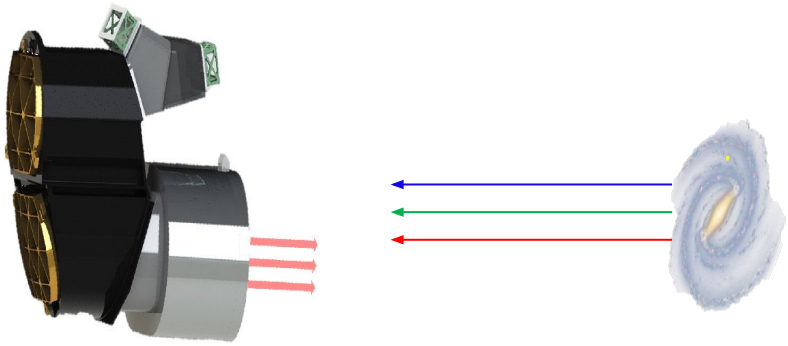
IVOA Interop May 2025

Anastasia Laity (Caltech-IPAC/IRSA)

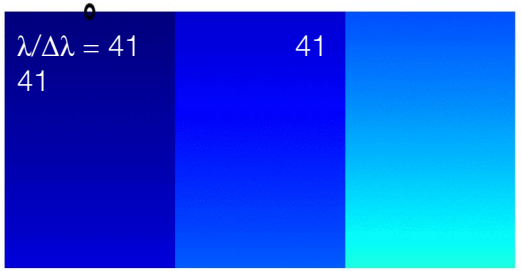
# SPHEREx IN A NUTSHELL



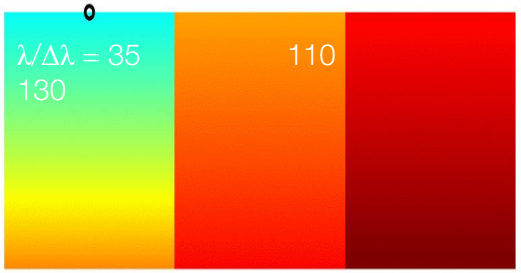

**Pointing modulates the wavelength of observation**

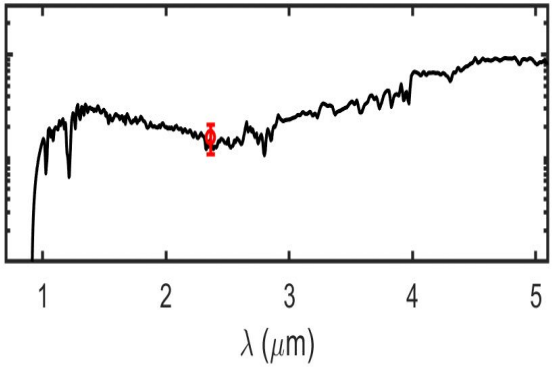


**Reflected by Dichroic**




**Transmitted by Dichroic**




A complete spectrum in 51 exposures
Each exposure takes 115s
1 complete spectrum every 6 months

# SPHEREx Data Products at IRSA

| | |
|---|---|
| **Quick Release Products:**<br>● **Calibrated spectral images**<br>● **Calibration files** | **Weekly releases starting Summer 2025** |
| **Spectrophotometry Tool:**<br>● **Operates on QR spectral images** | **Summer 2025** |
| Year 1&2 Data Releases<br>● Level 2 spectral images<br>● Calibration files<br>● 102 all-sky Spectral Data Cubes<br>● High Reliability Source Catalogs (spectra) | 2026 and 2027 |
| Additional Science Tools:<br>● Spectral image cutout<br>● Source discovery<br>● Custom Mosaic<br>● Spectral Cube cutout | Exact dates TBD |
| Legacy data products:<br>● Deep field mosaics<br>● Galaxy catalog<br>● Stellar type/ice column density catalog<br>● Stellar, galaxy cluster and solar system catalogs | mid-2028 |

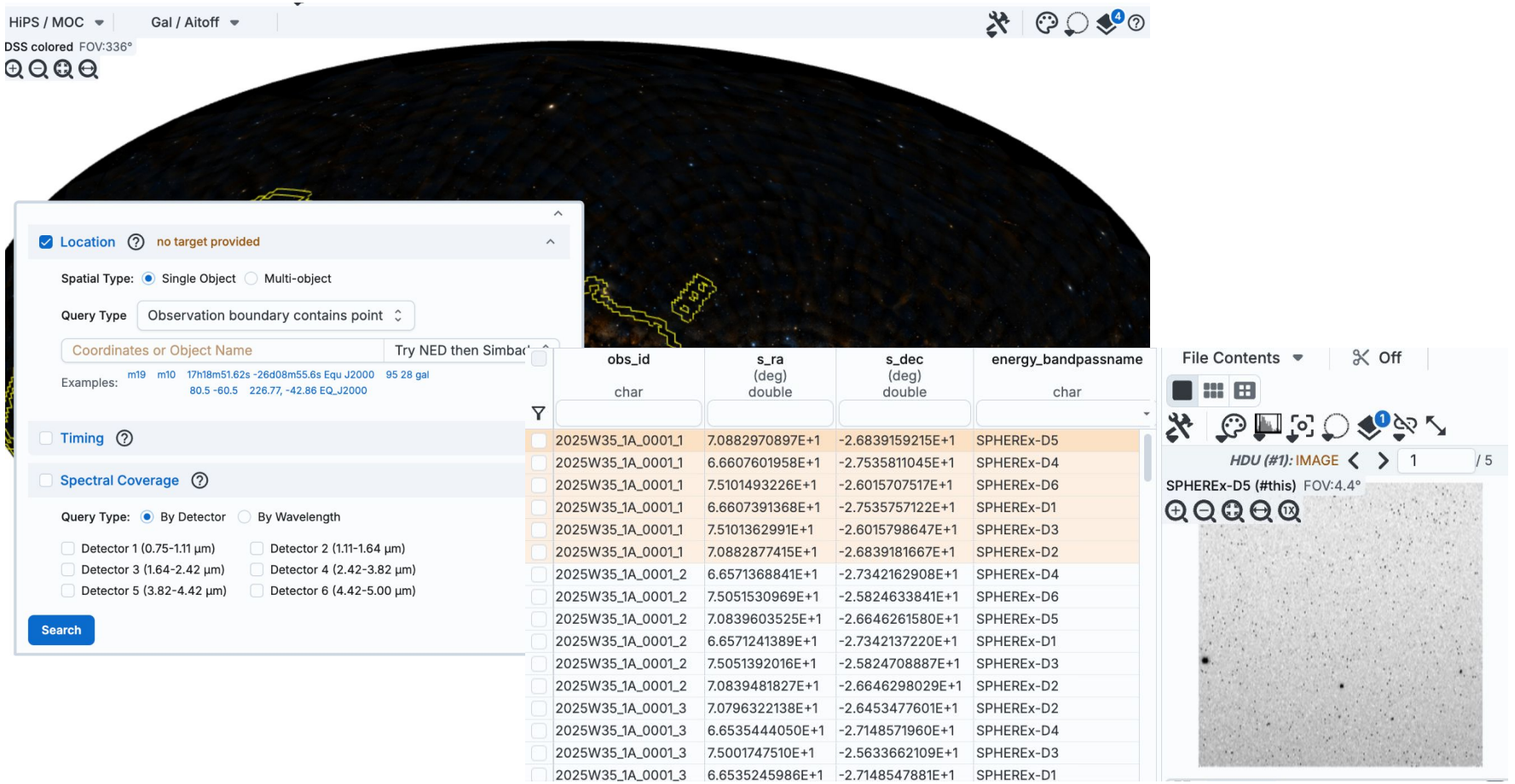
# SPHEREx at IRSA - Summer 2025

- Level 2 spectral images
  - Retrieve via: SPHEREx portal, SIA, ObsTAP
- Calibration files
  - Retrieve via: SIA, ObsTAP (first release)
  - SPHEREx portal via datalink by Fall 2025
- Storage:
  - On-prem and in cloud
  - CAOM data model with ObsCore view
- Spectrophotometry tool
  - Asynchronous
  - UWS protocol - callable via SPHEREx portal or directly

# Backend made easier - ObsCore / ObsTAP

- Primary data product search (calibrated level 2 spectral images) retrieved via ObsTAP using spherex.obscore (SPHEREx-only)

- SPHEREx datalink service returns service descriptors for cutouts, will include associated calibration files in incremental release



# Backend made easier - CAOM Collaboration

- Populated CAOM/ObsCore models together
  - SSDC and IRSA developers and scientists coordinated so that content made sense at all points in the IRSA stack
  - GUI - ObsCore view - CAOM metadata: need to think through the implications of low-level decisions.  (e.g. obs_publisher_did components)
- SPHEREx delivery mechanism populates CAOM tables
  - IRSA doesn't have to do any additional conversion; just ingest
  - First time working with project to deliver already-converted CAOM

**Still some challenges!**

"CAOM / ObsCore mapping is well known" != "IRSA CAOM implementation / ObsCore will be easy"

- Lots of "fiddly bits": formatting, spatial representation, etc.
- Calibration association
  - IRSA previously used plane/artifact association to relate science/calibration files (instead of maintaining provenance mode)
  - Not technically feasible for SPHEREx association model - ask me on a break!
  - For SPHEREx, ingesting calibration files as their own entities (separate collection). Ingesting provenance associations, will add DataLink support for it later this year

# Service Descriptor Syntax

**REST-style syntax vs service-descriptor-friendly:**

- legacy IRSA cutout service uses REST-style: path to file is part of URL

  `http://irsa.ipac.caltech.eduibe/cutout/spherex/lvf/20240927/level2/6/level2_2025W52_2W_0658_3D6_spx_P91_090124_level2.fits`**`?ra=164.28&dec=41.28&size=0.01`**

- DataLink table with multiple files needs service descriptor that takes input path as a parameter:

  `http://irsa.ipac.caltech.eduibe/cutout`**`?path=spherex/lvf/20240927/level2/6/level2_2025W52_2W_0658_3D6_spx_P91_090124_level2.fits&ra=164.28&dec=41.28&size=0.01`**

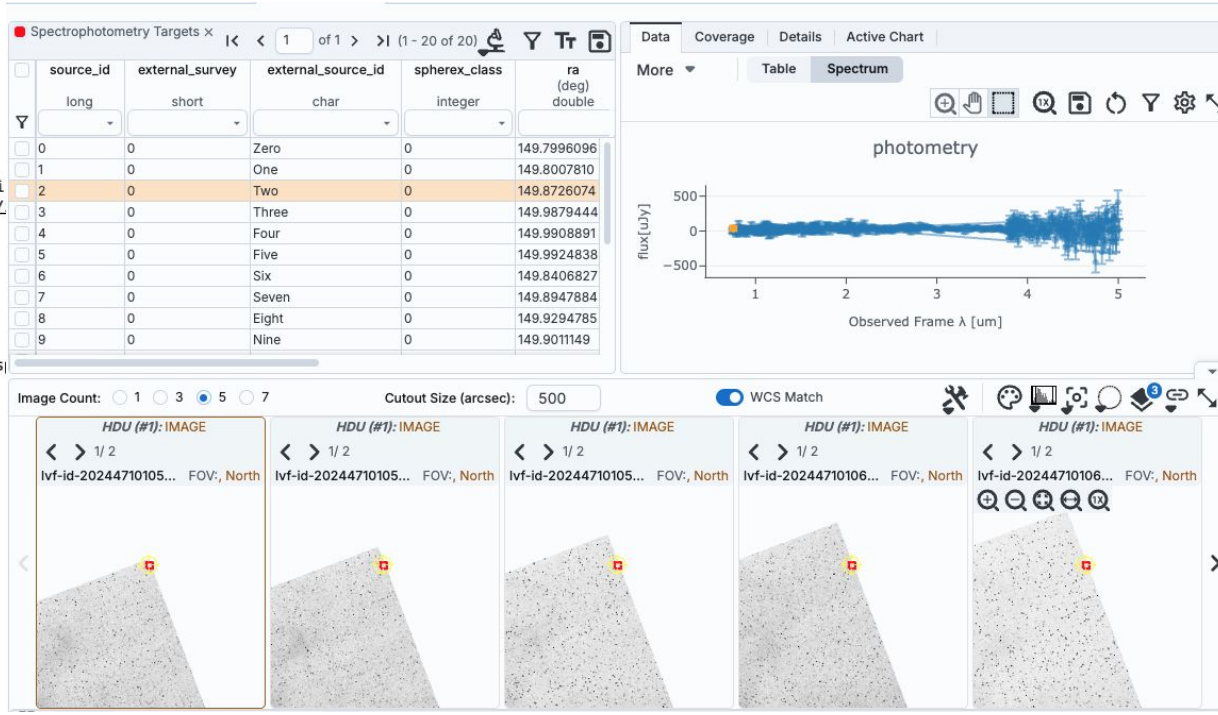
- *Result: IRSA updating cutout service to support both syntaxes, and revisiting how we design new APIs*

# Service Descriptors Syntax: Array Values

## Service Descriptors and Arrays of Input Parameters

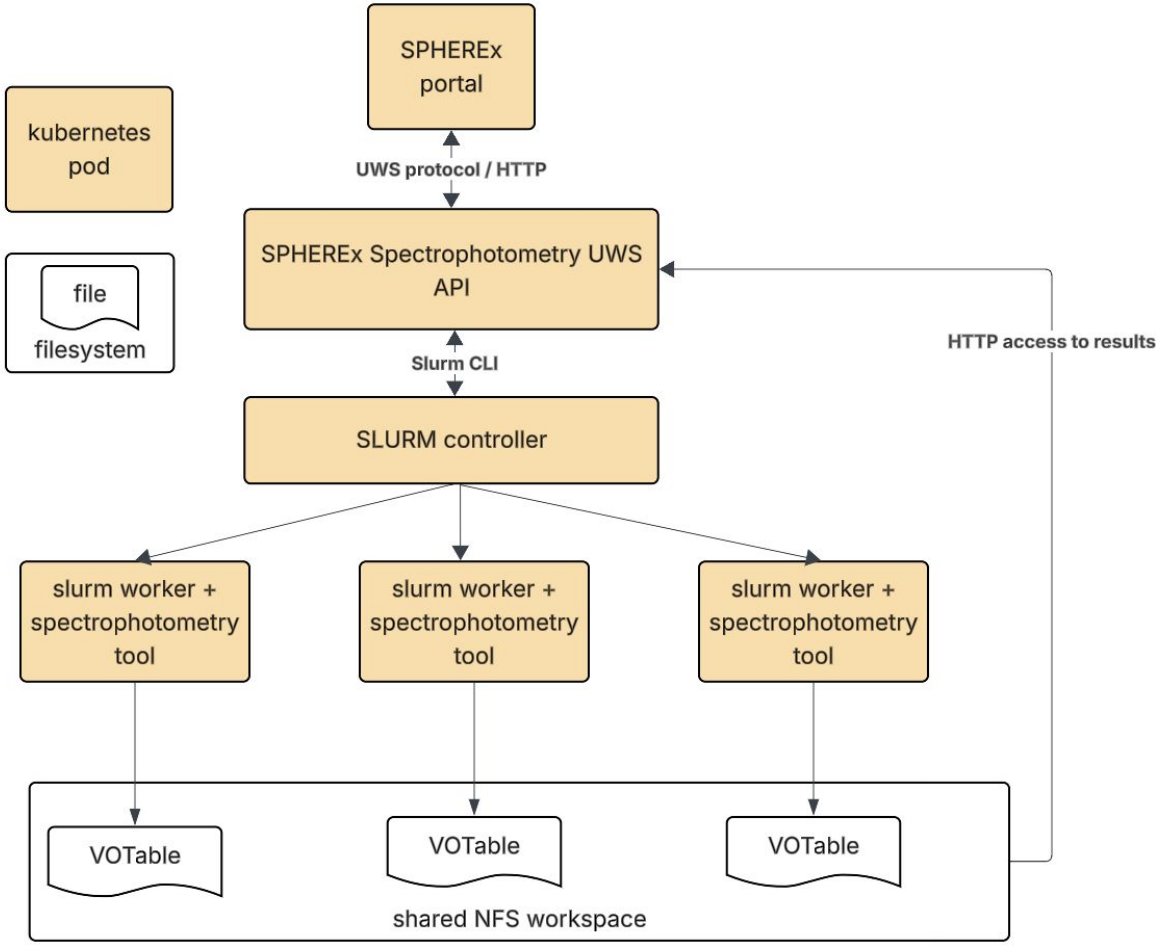
- Cutouts shown for all the input LVF images used to generate the spectra
- Spectra = 1 row; source images = array-value column

```
<RESOURCE type="meta" utype="adhoc:service">
    <DESCRIPTION>Simple LVF-cutout service</DESCRIPTION>
    <PARAM name="standardID" datatype="char" arraysize="*" value="ivo://i
    <PARAM arraysize="*" datatype="char" name="accessURL" value="https://
...
        <PARAM name="ID" datatype="char" arraysize="*">
            <DESCRIPTION>ID of LVF-image</DESCRIPTION>
        </PARAM>
    </GROUP>
</RESOURCE>

<!--  utype="ipac:MultiSpectrum" To indicate document contains multiple s
<RESOURCE type="results" utype="ipac:MultiSpectrum">
    <TABLE utype="ipac:MultiSpectrum">

        <!--  To indicate the data is packed in an array    -->
        <GROUP ID="photometry" utype="ipac:Spectrum.ArrayData">
            <GROUP utype="spec:Spectrum.Data.SpectralAxis">
                <FIELDref ref="lambda"/>
...
            </GROUP>
            <GROUP utype="spec:Spectrum.Data.FluxAxis">
                <FIELDref ref="flux"/>
...
            </GROUP>
            <!-- additional related columns -->
...
            <FIELDref ref="lvf_id" />
        </GROUP>
...
        <FIELD ID="flux" arraysize="*" datatype="float" name="flux" unit="uJy" utype="spec:Spectrum.Data.FluxAxis.Value"/>
...
        <FIELD ID="lvf_id" arraysize="*" datatype="long" name="lvf_id"/>
...
        <DATA>
            <TABLEDATA>
                <TR>
...
                    <TD>0.75375575 0.7713588 0.789379 0.80782604 0.82670957 0.8460397 0.8658268 1.0763234 1.101533 1.1273341 0.88608795 0.90682065
...
                    <TD>32.12902 65.24915 59.812084 58.238712 65.74049 119.70626 69.18804 127.732925 152.6282 98.83842 50.20299 88.4479 99.74918 7
...
                    <TD>202447101059121 202447101059131 202447101059141 202447101061011 202447101061021 202447101061031 202447101061041 2024481010
...
```



# UWS and Job Scheduling for Spectrophotometry

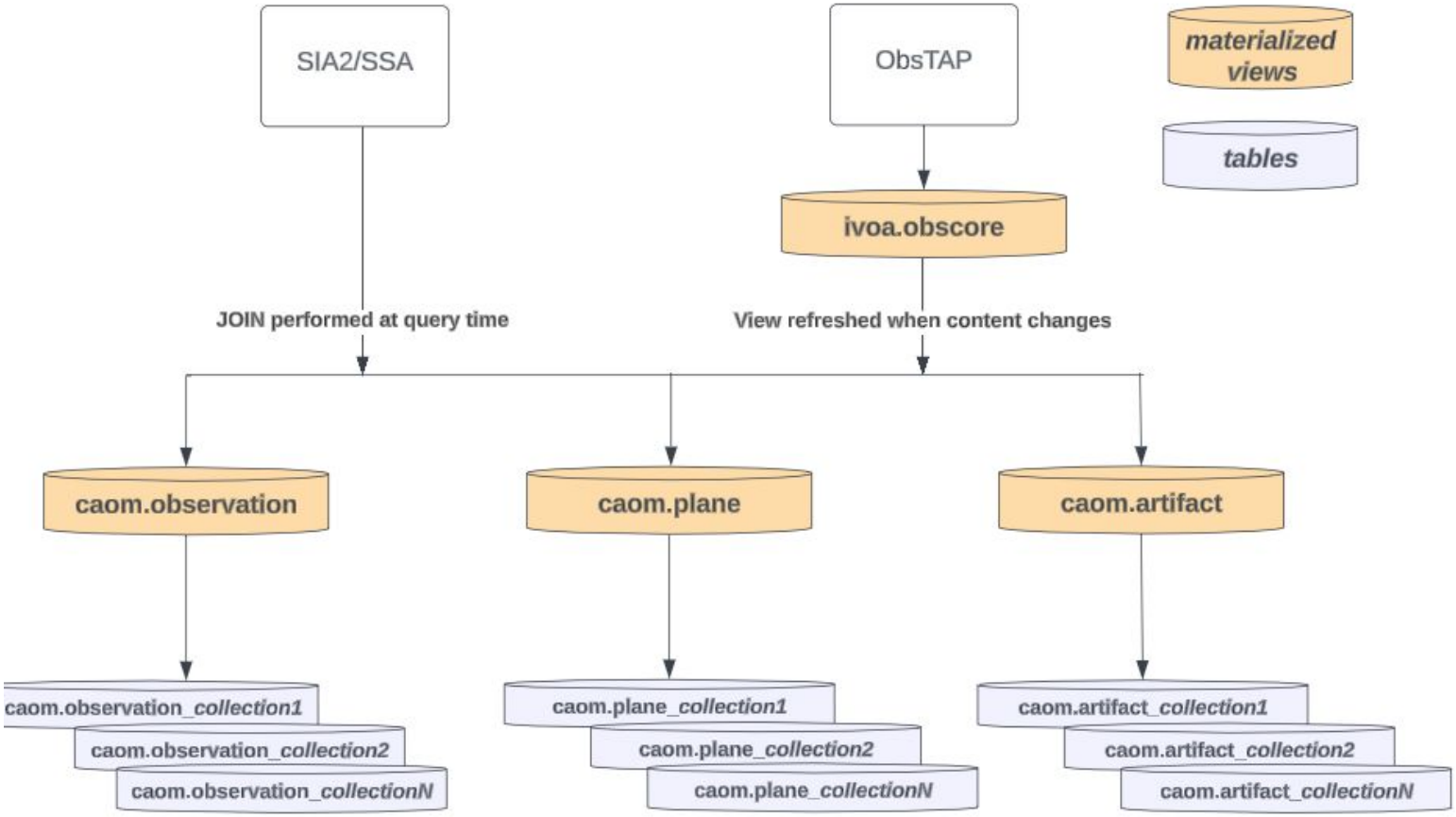
- Made several efforts to re-use existing general UWS tools
- Wound up developing service specific to tool (although UWS protocol layer + slurm configuration will be re-usable at IRSA)



# Infrastructure Challenge: SPHEREx weekly deliveries

**How we've managed CAOM/ObsCore til now**

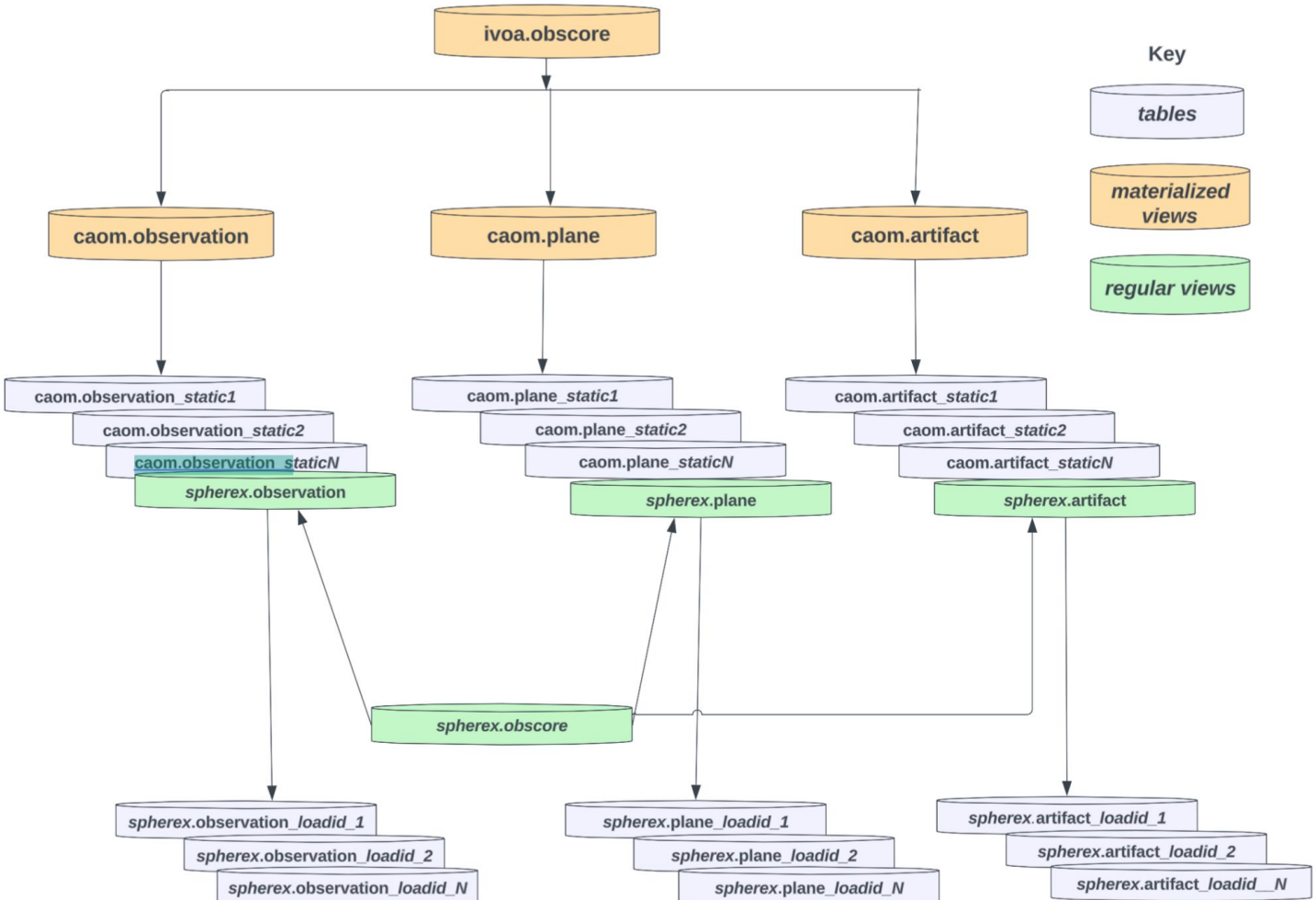
- Refreshed on demand (not more than once a month)
- Takes most of a day



# Infrastructure Challenge: SPHEREx weekly deliveries

**Supporting weekly updates for SPHEREx (and eventually others?):**

- GUI will use spherex schema (spherex.obscore) - get updates immediately
- SIA / ObsTAP will have some latency while main MVs rebuild

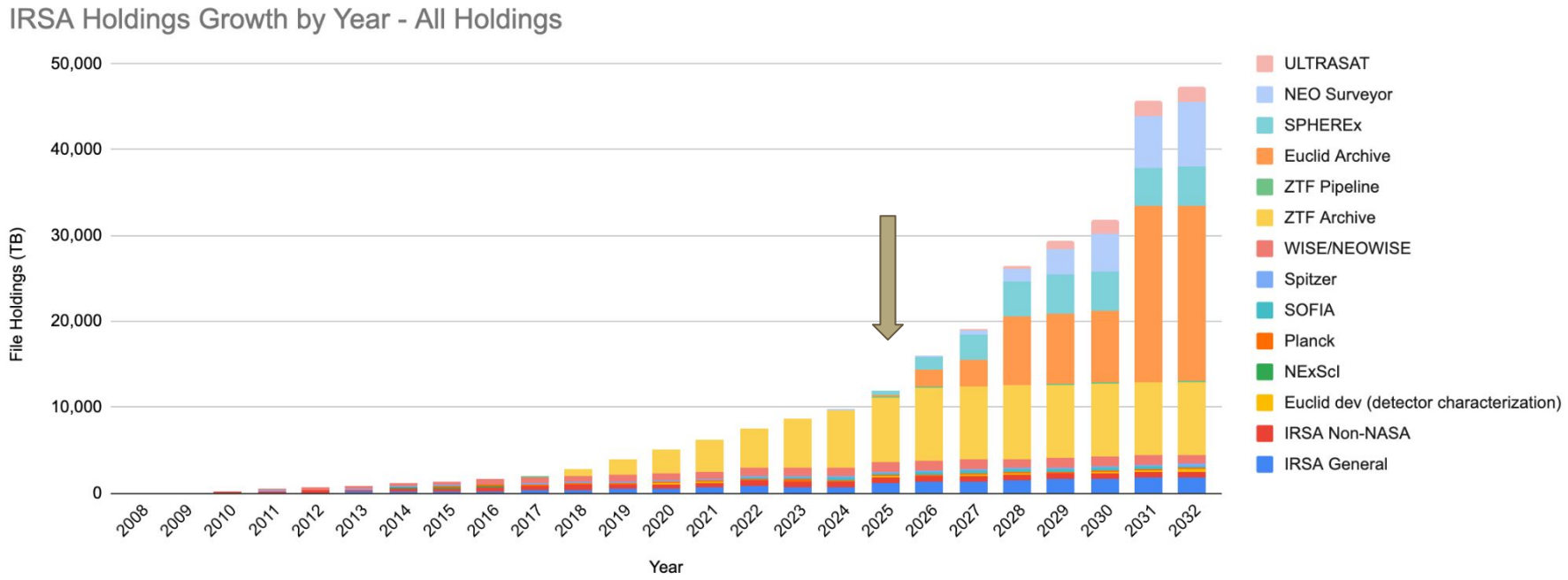


# Long-term challenge: overarching views + massive data holdings

**Ideal: use ObsTAP for data discovery as well as data retrieval**

**Reality: combining all holdings into one view presents latency/performance challenges that continue to grow with our holdings**

Revisit traditional IRSA model?  1) coverage check that lets users see what collections cover their area of interest, then 2) search those collections directly



# Takeaways

- ObsTAP was a good fit for SPHEREx spectral image search and considerably simplified IRSA's development of a search portal API
- Complexity of SPHEREx data presentation -> complicated service descriptor work; the need for things to "work as a service descriptor" is often a challenge
- Long-running SPHEREx tools needed to generate high-level products still required development of custom tools, but developing the layer that actually "speaks" UWS was simplified thanks to Josh Fraustro's UWS/openAPI work for P3T
- CAOM/ObsCore content management is complicated by weekly SPHEREx releases - and performance will be increasingly strained as our data size grows.

# Thank you!

IRSA / SSDC Contributers:

Rachel Akeson
Patrick Baghdasarian
Suela Buzi
Vandana Desai
Tamim Fatahi
Gregory Dubois-Felsmann
Tatiana Goldina
Steve Groom
Justin Howell
Yuna Kwon
Loi Ly
Trey Roby
Judith Silverman
Jaladh Singhal
Phani Valicheti
Angela Zhang