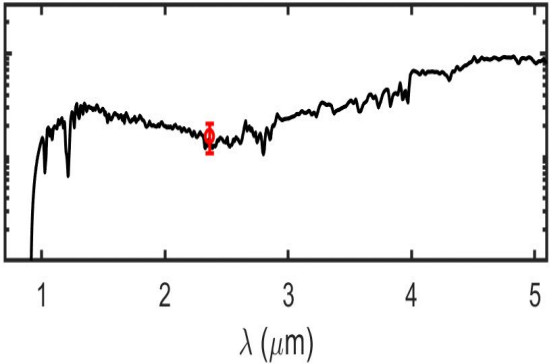
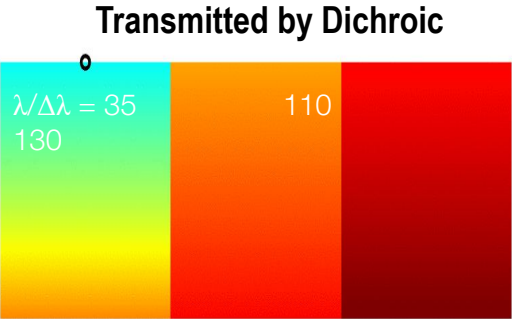
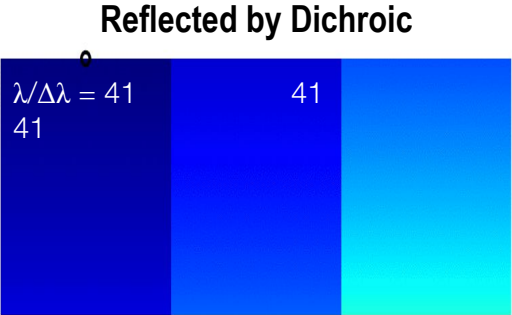
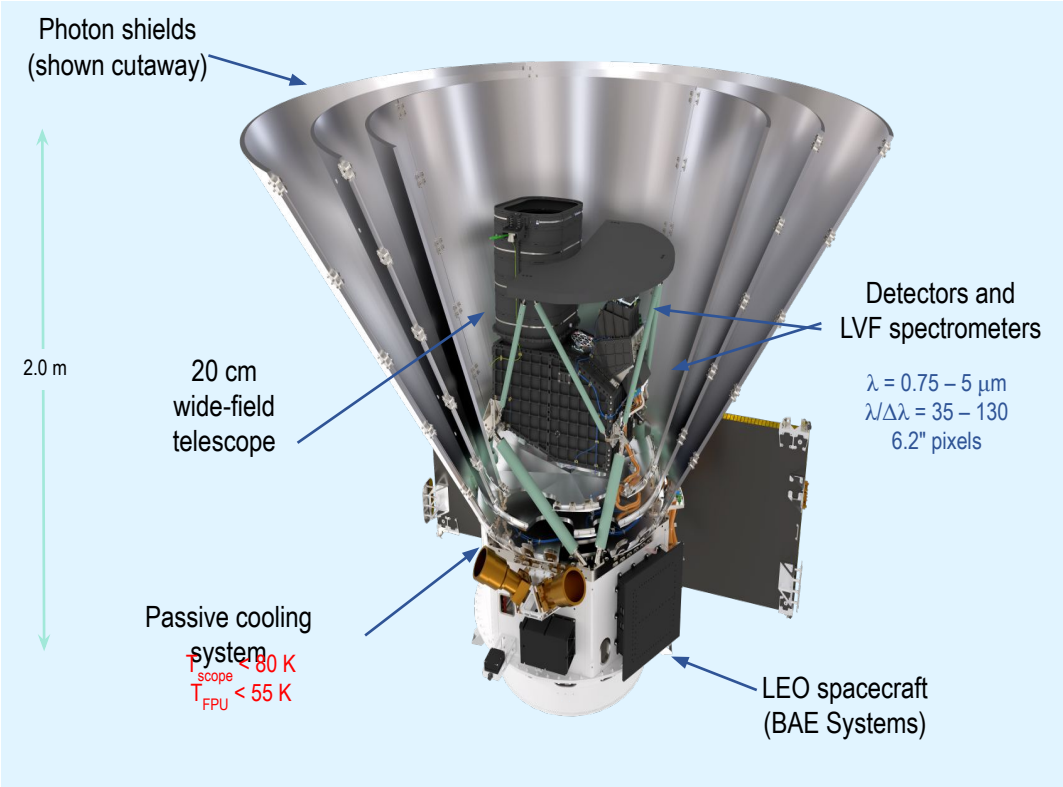

SPHEREx and the VO at IRSA

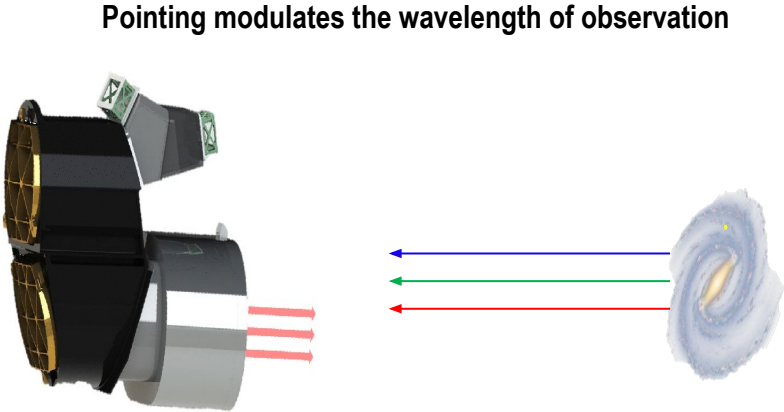
IVOA Interop May 2025

Anastasia Laity (Caltech-IPAC/IRSA)

SPHEREx IN A NUTSHELL



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



SPHEREx Data Products at IRSA

Quick Release Products: <ul style="list-style-type: none">• Calibrated spectral images• Calibration files	Weekly releases starting Summer 2025
Spectrophotometry Tool: <ul style="list-style-type: none">• Operates on QR spectral images	Summer 2025
Year 1&2 Data Releases <ul style="list-style-type: none">• Level 2 spectral images• Calibration files• 102 all-sky Spectral Data Cubes• High Reliability Source Catalogs (spectra)	2026 and 2027
Additional Science Tools: <ul style="list-style-type: none">• Spectral image cutout• Source discovery• Custom Mosaic• Spectral Cube cutout	Exact dates TBD
Legacy data products: <ul style="list-style-type: none">• Deep field mosaics• Galaxy catalog• Stellar type/ice column density catalog• 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

The screenshot displays the ObsTAP web interface. At the top, there are tabs for 'HiPS / MOC' and 'Gal / Aitoff'. Below these, a search bar is visible with the text 'DSS colored FOV:336°'. The main search area is a form with the following fields:

- Location** (checked): no target provided
- Spatial Type**: Single Object (selected), Multi-object
- Query Type**: Observation boundary contains point
- Coordinates or Object Name**: Try NED then Simbad
- Examples**: m19 m10 17h18m51.62s -26d08m55.6s Equ J2000 95 28 gal 80.5 -60.5 226.77, -42.86 EQ_J2000
- Timing** (unchecked)
- Spectral Coverage** (unchecked)
- Query Type**: By Detector (selected), By Wavelength
- Detectors**: Detector 1 (0.75-1.11 μm), Detector 2 (1.11-1.64 μm), Detector 3 (1.64-2.42 μm), Detector 4 (2.42-3.82 μm), Detector 5 (3.82-4.42 μm), Detector 6 (4.42-5.00 μm)

A **Search** button is located at the bottom left of the form. The search results are displayed in a table with the following columns:

obs_id	s_ra (deg)	s_dec (deg)	energy_bandpassname
char	double	double	char
<input type="checkbox"/> 2025W35_1A_0001_1	7.0882970897E+1	-2.6839159215E+1	SPHEREx-D5
<input type="checkbox"/> 2025W35_1A_0001_1	6.6607601958E+1	-2.7535811045E+1	SPHEREx-D4
<input type="checkbox"/> 2025W35_1A_0001_1	7.5101493226E+1	-2.6015707517E+1	SPHEREx-D6
<input type="checkbox"/> 2025W35_1A_0001_1	6.6607391368E+1	-2.7535757122E+1	SPHEREx-D1
<input type="checkbox"/> 2025W35_1A_0001_1	7.5101362991E+1	-2.6015798647E+1	SPHEREx-D3
<input type="checkbox"/> 2025W35_1A_0001_1	7.0882877415E+1	-2.6839181667E+1	SPHEREx-D2
<input type="checkbox"/> 2025W35_1A_0001_2	6.6571368841E+1	-2.7342162908E+1	SPHEREx-D4
<input type="checkbox"/> 2025W35_1A_0001_2	7.5051530969E+1	-2.5824633841E+1	SPHEREx-D6
<input type="checkbox"/> 2025W35_1A_0001_2	7.0839603525E+1	-2.6646261580E+1	SPHEREx-D5
<input type="checkbox"/> 2025W35_1A_0001_2	6.6571241389E+1	-2.7342137220E+1	SPHEREx-D1
<input type="checkbox"/> 2025W35_1A_0001_2	7.5051392016E+1	-2.5824708887E+1	SPHEREx-D3
<input type="checkbox"/> 2025W35_1A_0001_2	7.0839481827E+1	-2.6646298029E+1	SPHEREx-D2
<input type="checkbox"/> 2025W35_1A_0001_3	7.0796322138E+1	-2.6453477601E+1	SPHEREx-D2
<input type="checkbox"/> 2025W35_1A_0001_3	6.6535444050E+1	-2.7148571960E+1	SPHEREx-D4
<input type="checkbox"/> 2025W35_1A_0001_3	7.5001747510E+1	-2.5633662109E+1	SPHEREx-D3
<input type="checkbox"/> 2025W35_1A_0001_3	6.6535245986E+1	-2.7148547881E+1	SPHEREx-D1

On the right side, there is a **File Contents** panel with a search bar and a list of files. The selected file is **HDU (#1): IMAGE**. Below it, a preview of the data cutout is shown, labeled **SPHEREx-D5 (#this) FOV:4.4°**.

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_id 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="ad hoc:service">
  <DESCRIPTION>Simple LVF-cutout service</DESCRIPTION>
  <PARAM name="standardID" datatype="char" arraysize="*" value="ivo://i
...
  <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"/>
```

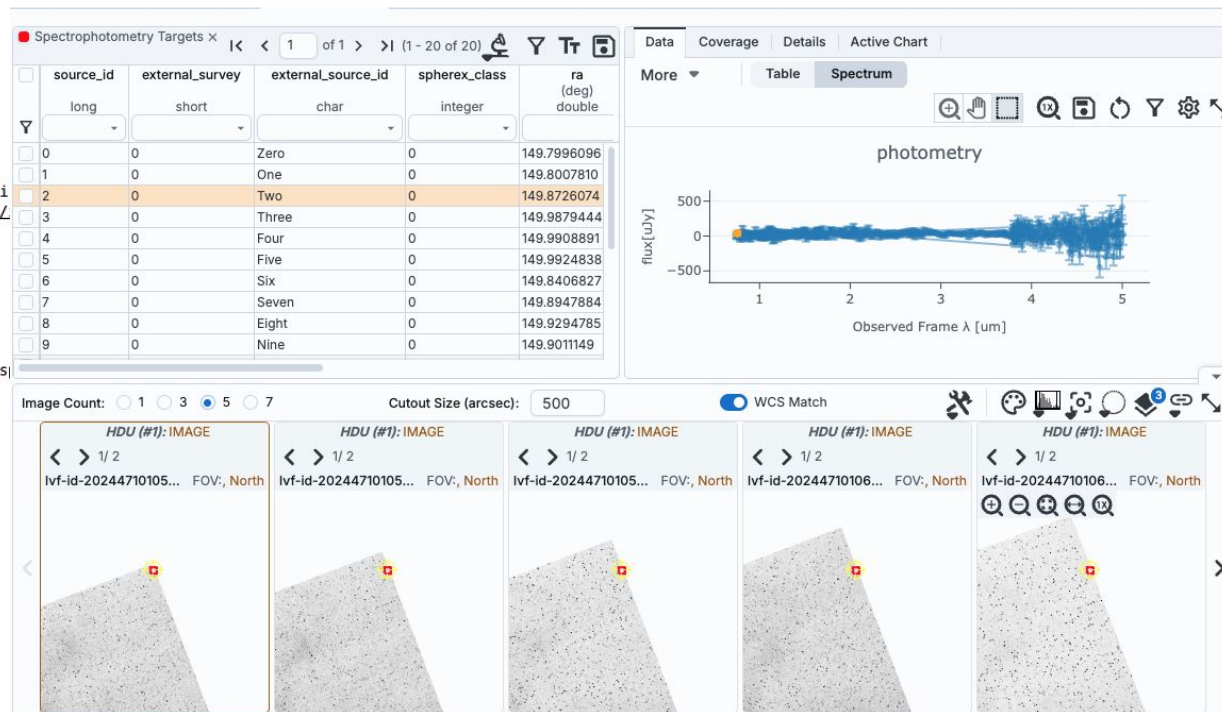
```
<FIELD ID="lvf_id" arraysize="*" datatype="long" name="lvf_id"/>
```

```
<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
```

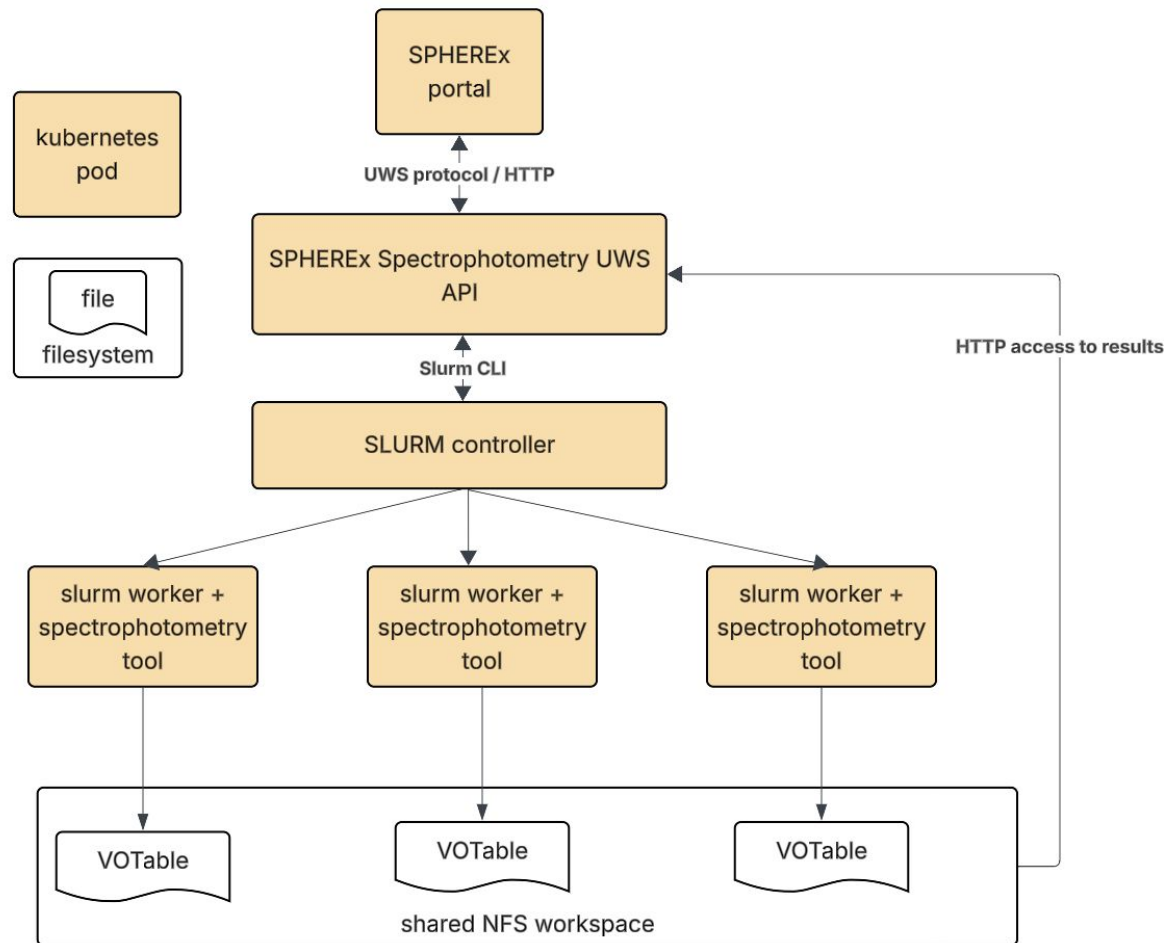
```
      <TD>32.12902 65.24915 59.812084 58.238712 65.74049 119.70626 69.18804 127.732925 152.6282 98.83842 50.20299 88.4479 99.74918 7
```

```
      <TD>202447101059121 202447101059131 202447101059141 202447101061011 202447101061021 202447101061031 202447101061041 2024481010
```



UWS and Job Scheduling for Spectrophotometry

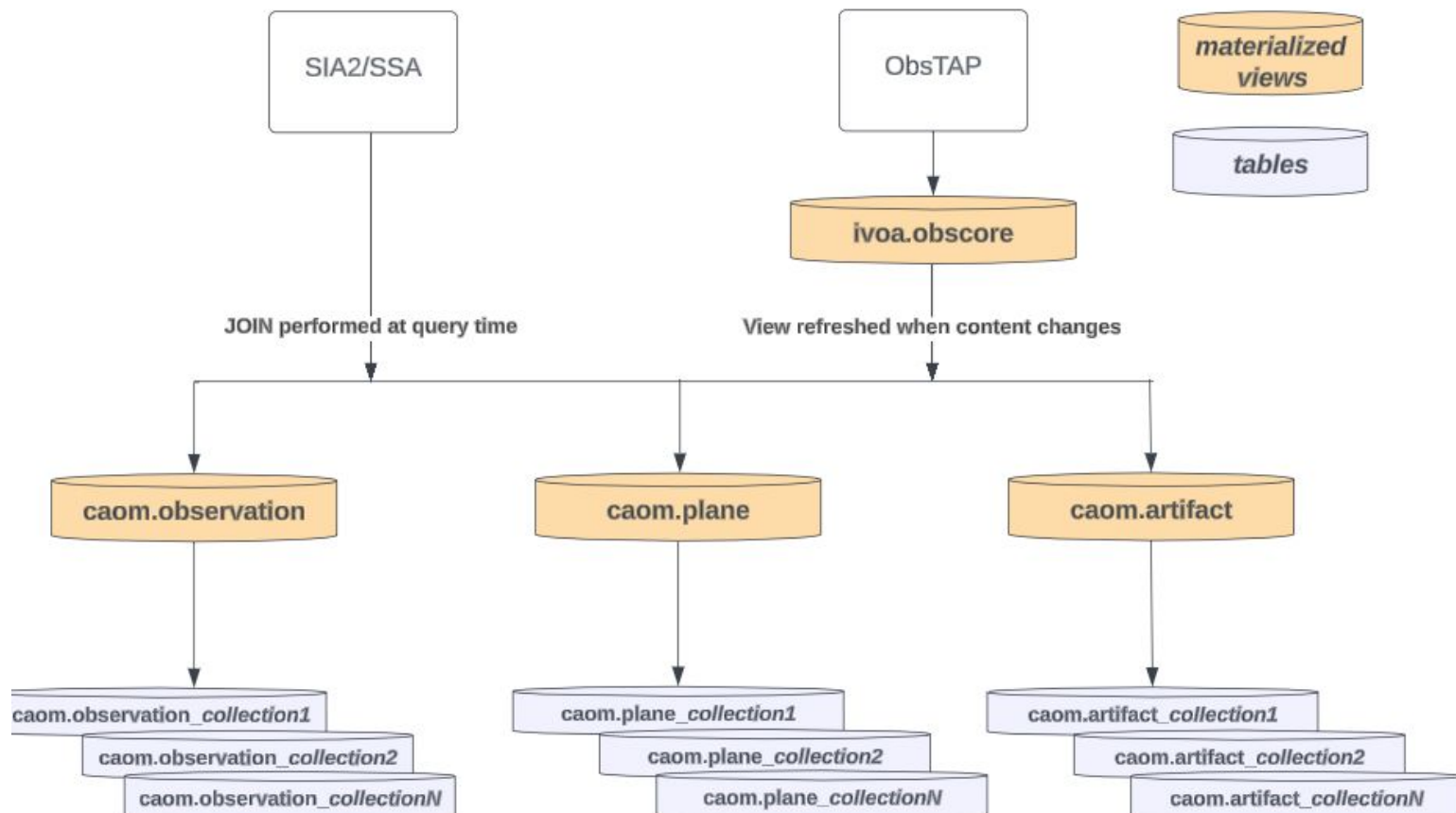
- Made several efforts to re-use existing general UWS tools
- Wound up developing service specific to tool (although UWS protocol layer + slurm configuration will be re-usable at IRSA)



Infrastructure Challenge: SPHEREx weekly deliveries

How we've managed CAOM/ObsCore til now

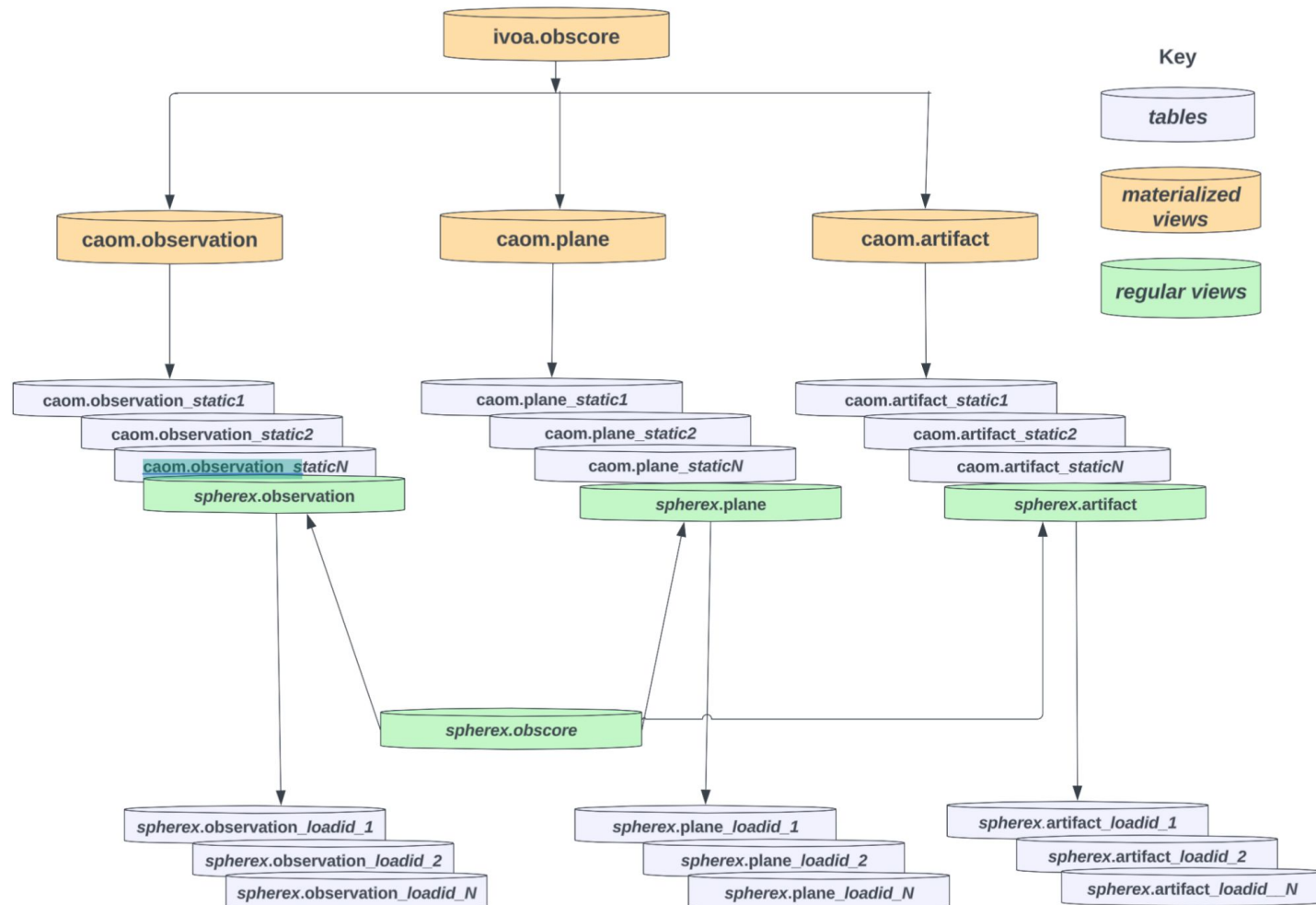
- Refreshed on demand (not more than once a month)
- Takes most of a day



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



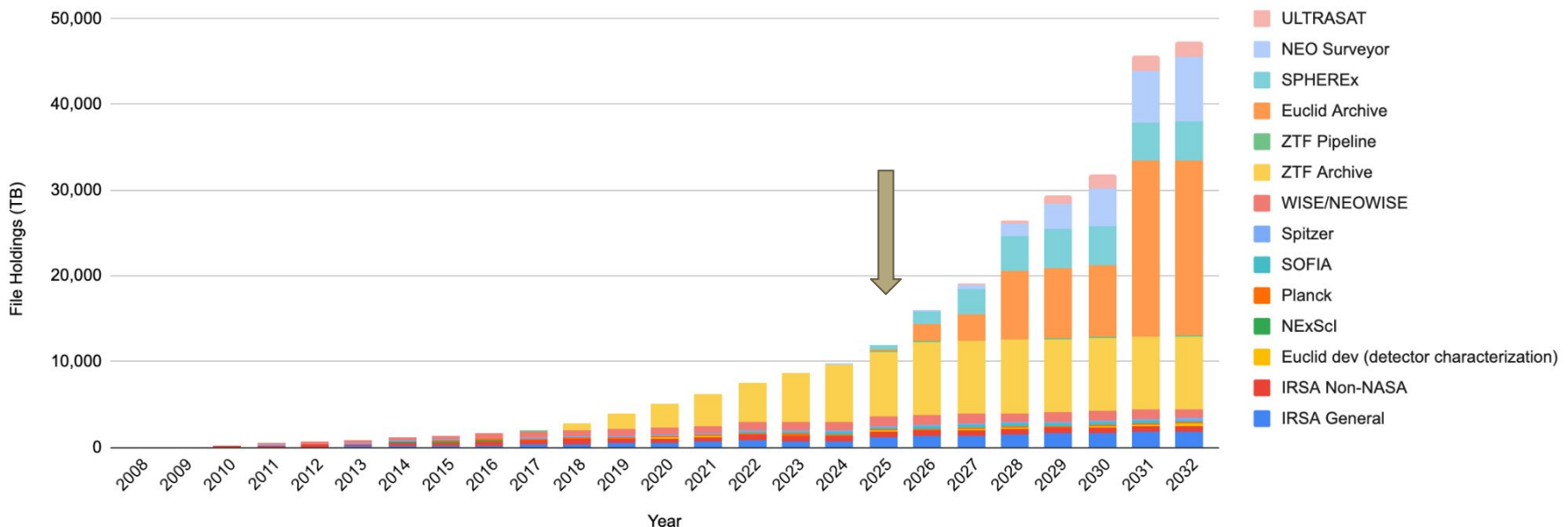
Long-term challenge: overarching views + massive data holdings

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

IRSA Holdings Growth by Year - All Holdings



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 Contributors:

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