DAL at IRSA: 2024 Updates

IVOA Interop Nov 2024

Anastasia Laity (Caltech-IPAC/IRSA)

Since May 2024 interop….

May 2024:

- Sydney interop: "NAVO Archives: Adventures in ObsTAP & DataLink"
- IRSA status: draft ivoa.obscore; design phase for DataLink service

This talk: what's new at IRSA since May?

Sep 2024:

- IRSA released dedicated VO services for simulated data

Nov 2024:

- Internal testing:
  - general IRSA DataLink service for ObsTAP/SxA
  - ivoa.obscore: view of of CAOM tables; uses IRSA DataLink service
  - SSA and SIA2 using DataLink

Serving Simulated Data in VO Services (1)

Need to make it difficult for users to confuse simulated data with real data

Scenario 1: user performs ObsTAP query

- NOT including the columns that clearly identify data as simulated
- Now simulated content unknowingly "mixed in" with real data

IRSA solution: serve data through separately registered services

- /simulated/SxA
- /simulated/TAP
  - Simulated content **only accessible** through these special endpoints
  - ivoa.obscore mapped to hidden ivoa.simulated_obscore - cannot be accessed via "normal" /TAP

Serving Simulated Data in VO Services (2)

Scenario 2: Client queries /simulated/TAP or /simulated/SxA on date X

- At date X + Y, they reopen that table that's been sitting on their desktop
- What if they forgot it was from simulated services?

IRSA Solutions:

- Additional column `irsa_provenance` **always included** in simulated endpoint results: value="`SIMULATED`"
- `obs_collection` has "`simulated_`" prefix
- Encourage inclusion of "`simulated`" in other columns
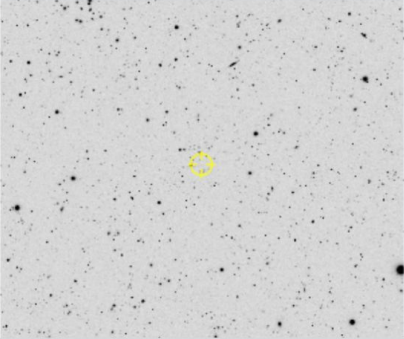
Simulated Data at IRSA - Example

https://irsa.ipac.caltech.edu/simulated/SIA?collection=simulated_lsstdesc2021

IRSA ObsTAP / DataLink - Examples

- **ObsTAP/SIA2:**
  - Each row = **"observation"**: group of artifacts/services = **caom.plane row**
  - `access_url` : URL to call DataLink service

calib_level	access_url	s_ra (deg)	s_dec (deg)	em_min (m)	em_max (m)	facility_name
short	char	double	double	double	double	char
2	https://irsawebdev2.ipac.caltech.edu/datalink/links?ID=ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1	294.1090167097	-21.4132454680	0.0000031500	0.0000037800	Wide-field Infrared Survey Explorer
2	https://irsawebdev2.ipac.caltech.edu/datalink/links?ID=ivo://irsa.ipac/neowiser/4r_49464r_113/4r_49464r_113_W1	294.0507430367	-21.3653135041	0.0000031500	0.0000037800	Wide-field Infrared Survey Explorer
2	https://irsawebdev2.ipac.caltech.edu/datalink/links?ID=ivo://irsa.ipac/neowiser/4r_49464r_113/4r_49464r_113_W2	294.0507225490	-21.3648042422	0.0000041000	0.0000051600	Wide-field Infrared Survey Explorer
2	https://irsawebdev2.ipac.caltech.edu/datalink/links?ID=ivo://irsa.ipac/neowiser/9r_20929r_114/9r_20929r_114_W1	294.0407953619	-21.3666352207	0.0000031500	0.0000037800	Wide-field Infrared Survey Explorer
2	https://irsawebdev2.ipac.caltech.edu/datalink/links?ID=ivo://irsa.ipac/neowiser/9r_20929r_114/9r_20929r_114_W2	294.0408165398	-21.3671417964	0.0000041000	0.0000051600	Wide-field Infrared Survey Explorer
2	https://irsawebdev2.ipac.caltech.edu/datalink/links?ID=ivo://irsa.ipac/neowiser/7r_32147r_139/7r_32147r_139_W1	294.0115988056	-21.4634166333	0.0000031500	0.0000037800	Wide-field Infrared Survey Explorer
2	https://irsawebdev2.ipac.caltech.edu/datalink/links?ID=ivo://irsa.ipac/neowiser/7r_32147r_139/7r_32147r_139_W2	294.0116225884	-21.4639184981	0.0000041000	0.0000051600	Wide-field Infrared Survey Explorer

- **DataLink Service:**
  - `https://irsa.ipac.caltech.edu/datalink/links?ID=`**obs_publisher_did**
  - Each row = **caom.artifact row** (or service descriptor reference)

id	access_url	service_def	semantics	description	content_type
char	char	char	char	char	char
ivo://irsa.ipac/neowiser/1a_75	https://irsa.ipac.caltech		#this	ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1 science image	image/fits
ivo://irsa.ipac/neowiser/1a_75		ibe-cutout	#cutout	ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1 science image	image/fits
ivo://irsa.ipac/neowiser/1a_75	https://irsa.ipac.caltech		#auxiliary	ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1 auxiliary image	application/x-gzip
ivo://irsa.ipac/neowiser/1a_75		ibe-cutout	#cutout	ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1 auxiliary image	application/x-gzip
ivo://irsa.ipac/neowiser/1a_75	https://irsa.ipac.caltech		#noise	ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1 noise image	application/x-gzip
ivo://irsa.ipac/neowiser/1a_75		ibe-cutout	#cutout	ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1 noise image	application/x-gzip
ivo://irsa.ipac/neowiser/1a_75	https://irsa.ipac.caltech		#auxiliary	ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1 auxiliary image	text/plain
ivo://irsa.ipac/neowiser/1a_75	https://irsa.ipac.caltech		#auxiliary	ivo://irsa.ipac/neowiser/1a_75331a_056/1a_75331a_056_W1 auxiliary image	text/plain

ObsTAP/DataLink - Barriers to Public Release

- Performance issues with DataLink service
  - If we can't fix it with better indexing / data types...
  - Supply planeid directly as extra service input when called from ObsTAP/SxA?

$$ID=obs\_publisher\_did\&FASTID=hash$$

- SIA2 rollout plan
  - Clients currently expect 1 row per artifact
  - How can we smooth the transition to DataLink call?

- Internal testing/feedback cycle
  - We have some data model questions... (next slide)

Data Model Questions

How much does consistency of ObsCore content across archives matter to scientists?

- **obs_collection**: some providers use data model names that are only internally meaningful
- **calib_level:** can be disagreement on 3 v 4
- **dataproduct_type:** not everyone sticks to defined vocab

How can we improve mapping from CAOM to ObsCore?

- `description` (DataLink)
  - We currently resort to filename to capture file-specific content
  - Would prefer something more readable
- `dataproduct_subtype` (ObsCore)
  - if no obvious place in CAOM 2.5, might add IRSA-internal CAOM column

Bonus Slide: DataLink not just for ObsTAP/SIA

- Custom DataLink API for a future mission's GUI - needs to return products from **multiple** observations
- Inputs:
  - **OBJECTID=(mission object identifier)**
  - No **ID** input supplied
- Products in same group have same **ID=obs_publisher_did**

id	access_url	service_def	semantics	description
char	char	char	char	char
ivo://irsa.ipac/mission_spec/123/123456	url_to_123456_spectrum.fits		#this	123456 Co-added spectrum
ivo://irsa.ipac/mission_tile/123/G	url_to_123_G.fits		#counterpart	123 G band tile
ivo://irsa.ipac/mission_tile/123/G		sd-cutout	#cutout	123 G band tile cutout
ivo://irsa.ipac/mission_tile/123/G	url_to_123_G_RMS.fits		#noise	123 G band RMS
ivo://irsa.ipac/mission_tile/123/G	url_to_123_G_PSF.fits		#auxiliary	123 G band PSF
ivo://irsa.ipac/mission_obs/987/G-001	url_to_987_G_001.fits		#counterpart	987 G band dither 001
ivo://irsa.ipac/mission_obs/987/G-001		sd-cutout	#cutout	987 G band dither 001 cutout
ivo://irsa.ipac/mission_obs/987/G-002	url_to_987_G_002.fits		#counterpart	987 G band dither 002
ivo://irsa.ipac/mission_obs/987/G-002		sd-cutout	#cutout	987 G band dither 002 cutout

Thank you!

Contributors & consultants at IRSA:

Tim Burke

Vandana Desai

Gregory Dubois-Felsmann

Justin Howell

Loi Ly

Trey Roby

Judith Silverman

Scott Terek

Angela Zhang