Radio (and millimetric) Data in the VO : State of the art

F.Bonnarel (CDS)

acknowledges extensive collaborative work with DAL WG, DM WG, TDIG, DADI collaborators, Strabourg CDS and SVOM teams

CENTRE DE DONNÉE ASTRONOMIQUES DE STRASBOURG

Astronomy ESFRI & Research Infrastructure Cluster

IVOA

Summary

- I) multi-dimensional data in the VO
  - Examples of projects
  - Problems for interoperabilty
  - DataModels overview
  - Discovery solutions : ObsCore, SIAV2
  - Access Solutions : DataLink and SODA
- II) Client-side Interfaces
  - Aladin
  - TOPCAT
- III) ALMA Science Archive implementation
  - Available data in the archive
  - Archive ObsTAP Interface
- IV) ASKAP data : CASDA implementation

Multi-dimensional data

- Radio datacubes
- Xray cubes
- Time dimensions
- Polarisation
- Visibility data

Radio data Cubes

- ALMA
- LOFAR, ASKAP
- Towards SKA

→data discovery by selecting criteria on description attributes (metadata)

→data access : extracting useful data from the datasets

DataAccess

Radio Cubes are « huge »

→ full retrieval cumbersome

→ Cutouts

→ eventually regridding, reprocessing ?

→ need for detailed data (metadata) representation further than ObsCore (Cube DM, may come later)….

VO modelling

1) Obscore : metadata for discovery

obs_collection	unitless	String	Name of the data collection
obs_id	unitless	String	Observation ID
obs_publisher_did	unitless	String	Dataset identifier given by the publisher
access_url	unitless	String	URL used to access (download) dataset
access_format	unitless	String	File content format (see in App. BB.5.2)
access_estsize	kbyte	integer	Estimated size of dataset in kilo bytes
target_name	unitless	String	Astronomical object observed, if any
s_ra	deg	double	Central right ascension, ICRS
s_dec	deg	double	Central declination, ICRS
s_fov	deg	double	Diameter (bounds) of the covered region
s_region	unitless	String	Sky region covered by the data product (expressed in ICRS frame)
s_xel1	unitless	integer	Number of elements along the first spatial axis
s_xel2	unitless	integer	Number of elements along the second spatial axis
s_resolution	arcsec	double	Spatial resolution of data as FWHM
t_min	d	double	Start time in MJD
t_max	d	double	Stop time in MJD
t_exptime	s	double	Total exposure time
t_resolution	s	double	Temporal resolution FWHM
t_xel	unitless	integer	Number of elements along the time axis
em_min	m	double	Start in spectral coordinates
em_max	m	double	Stop in spectral coordinates
em_res_power	unitless	double	Spectral resolving power
em_xel	unitless	integer	Number of elements along the spectral axis

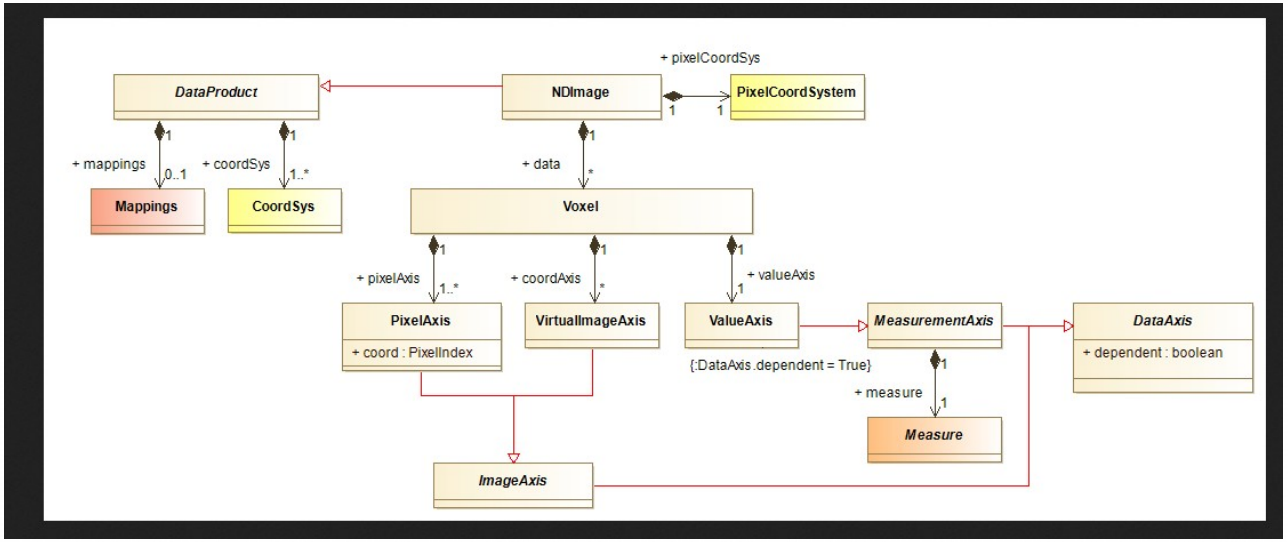
VO modelling

2) Other existing useful data models

- STC 1.0 : useful concepts, currently under strong revision (Coords, Measurements and Transformations)
- PhotDM : mostly for Optical astronomy
- Characterisation 1.0 : reused by other datamodels (such as ObsCore, Cube) under revision
- Spectral Line Data Model : From observation to Physics
- Simulation DataModels. Comparison with Simulations

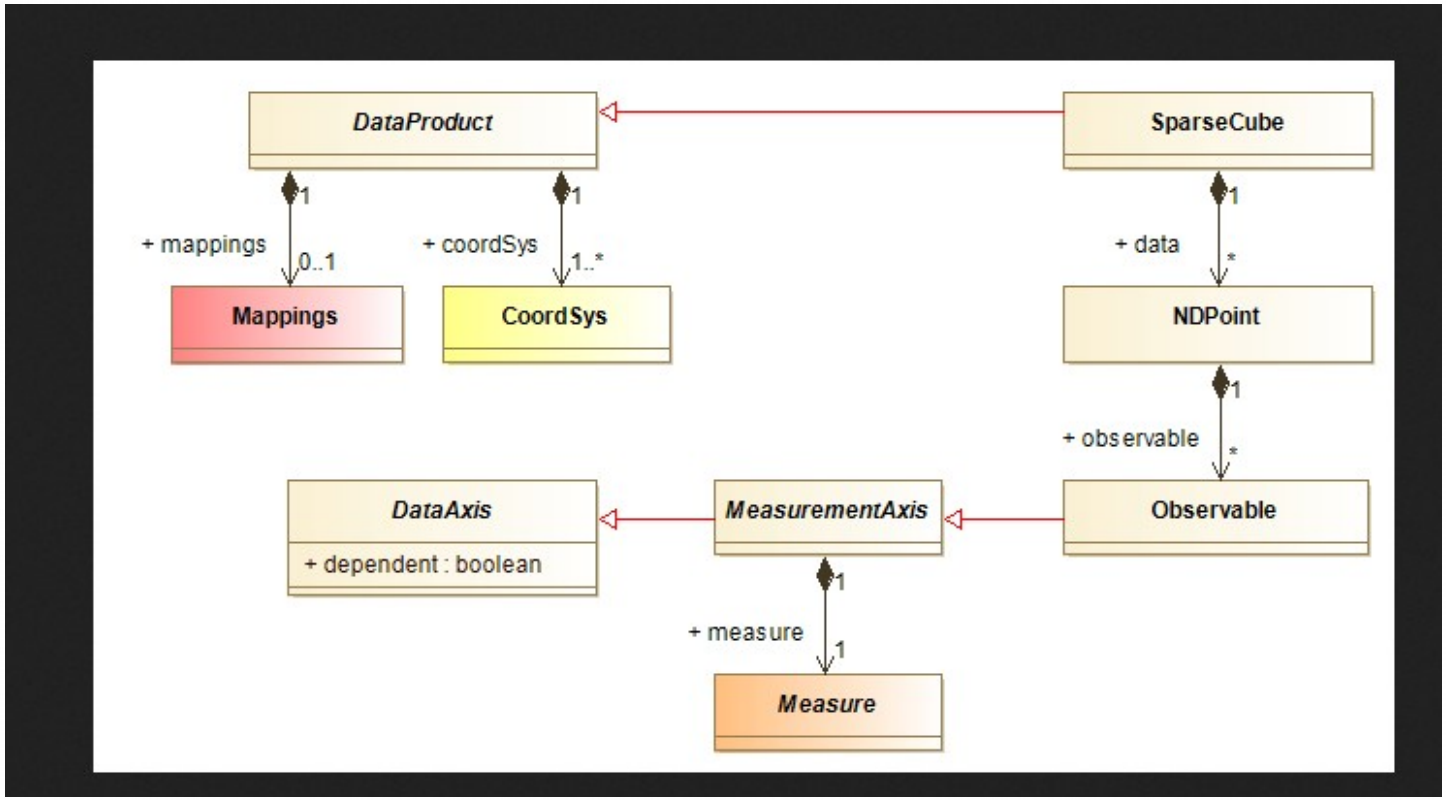
VO modelling

3) Cube datamodel (in progress) : ND Image



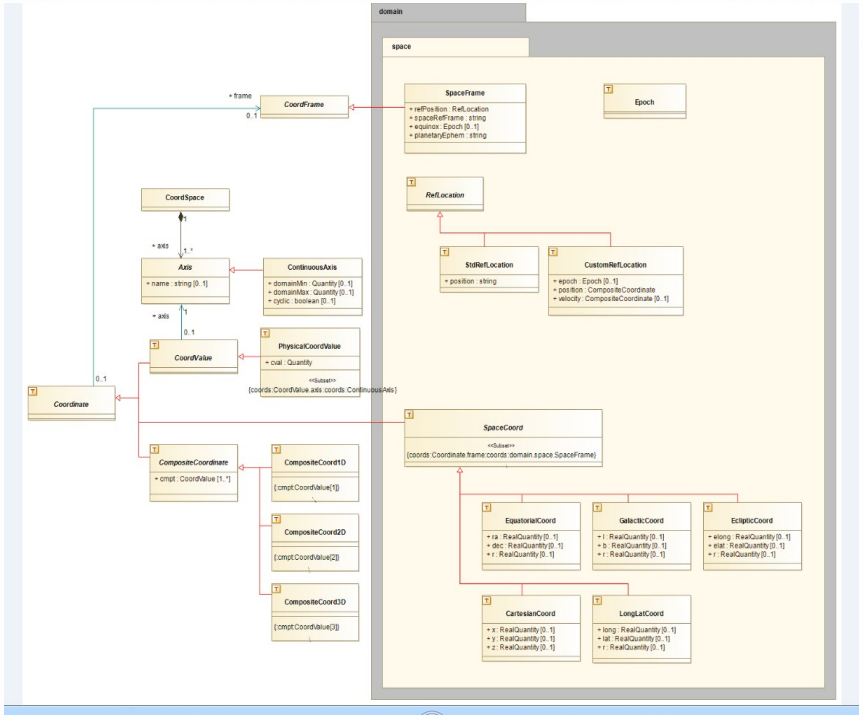
VO modelling

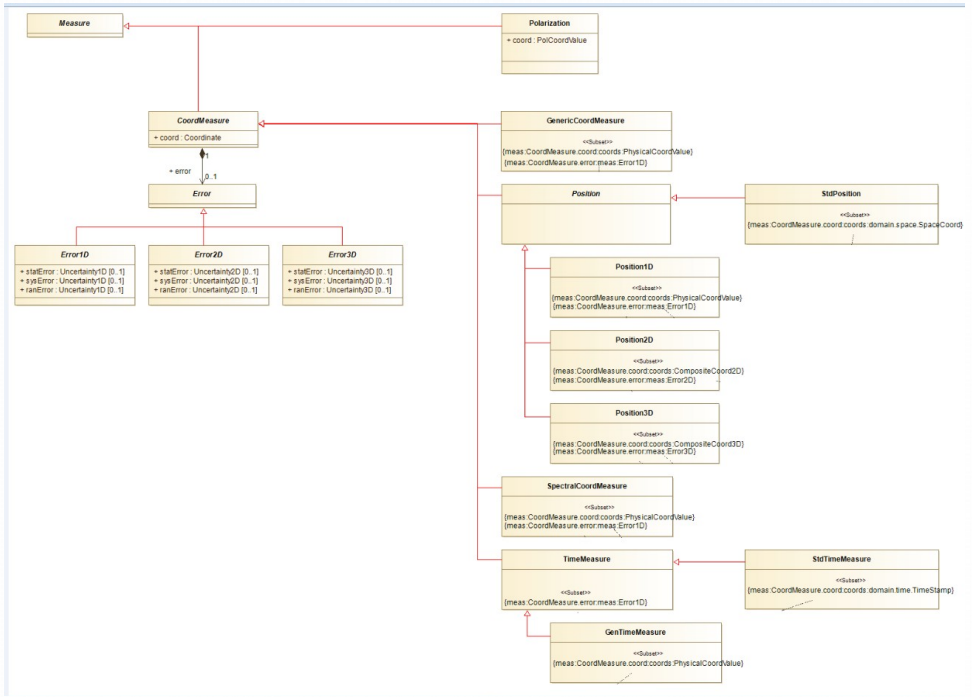
3 bis) Cube datamodel : sparse cube



VO datamodelling

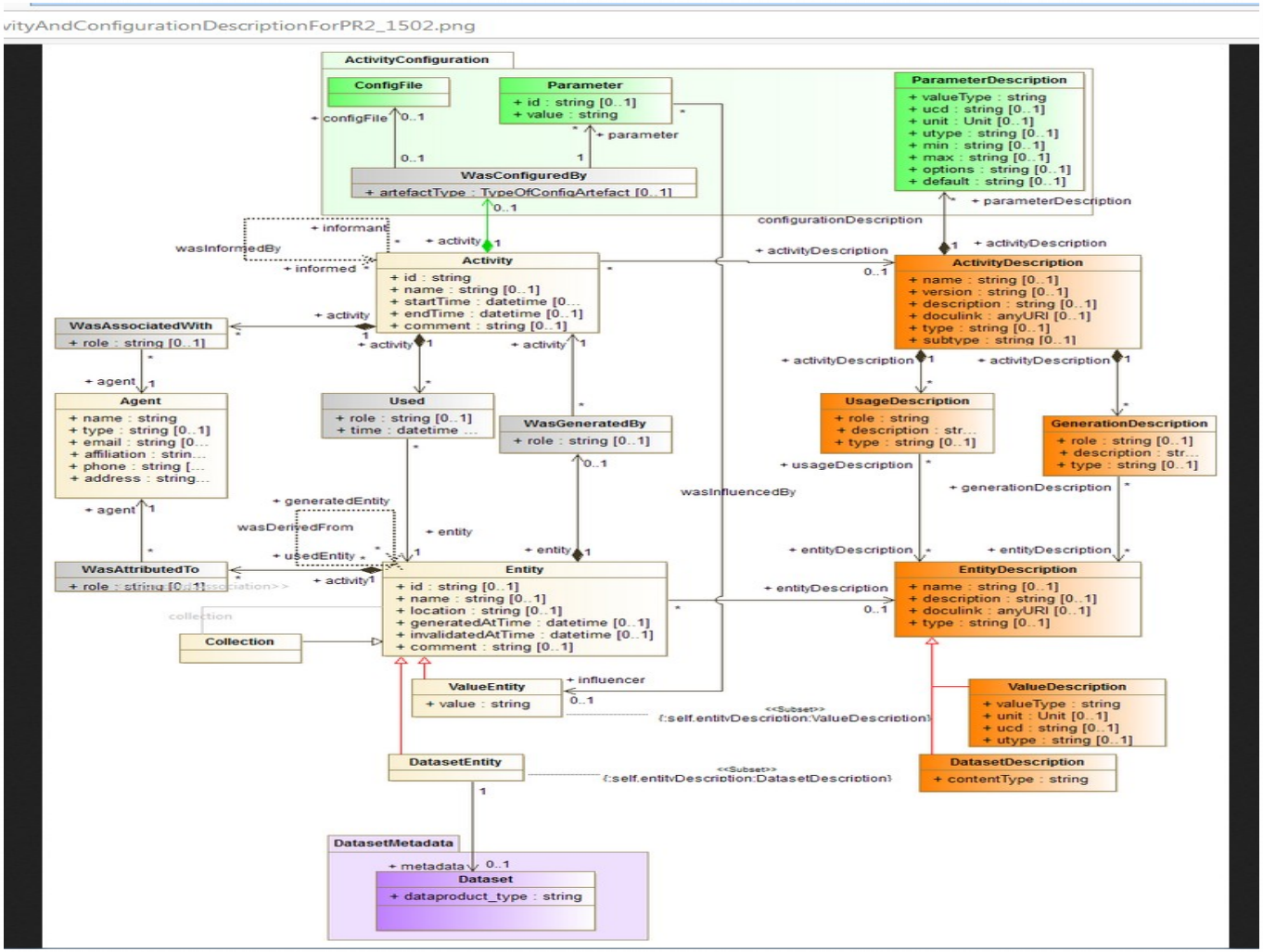
4) STC : Coords and Measurements (in progress)





VO datamodelling

5) Provenance (in progress)



VO access layer Solutions

- ObsTAP (serving ObsCORE table) : allows to select datasets on criteria such as dataproduct_type= cube and characterisation of axes
- SIAV2 : allows queries for images and cubes with PARAMETERS
  - such as DPTYPE = cube, POS = …. BAND = 0.0005 0.0006 TIME = 52618 53700 etc.
- Full Data Retrieval via access_reference field (if manageable)
  - Or access to DataLink …. and SODA (direct access to SODA also possible)
- Alternative : Discovery , Representation and View via HiPS for cubes

SIAV2 parameters

POS (spatial constraints)

BAND (spectral constraints)

TIME (time constraints)

POL (polarisation constraints)

SPATRES (spatial resolution constraints)

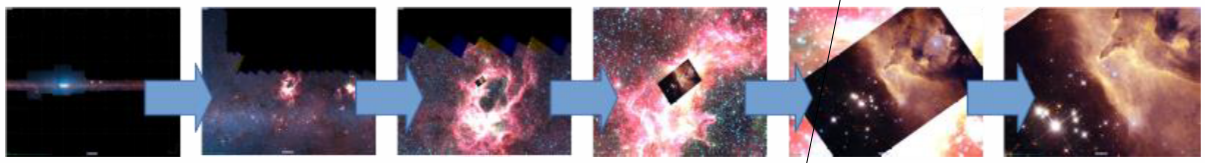
SPECRES (spectral resolution constraints)

COLLECTION (restrict to a data collection)

FACILITY (restrict to a facility)

HiPS ? What is it ?

- **Hi**erachical **P**rogressive **S**urvey
  "The more you zoom in on a particular area,the more details show up"
- Multi-resolution **HEALPix** data structure
- for **Images**, **Catalogues**, 3-dimensional data **cubes**, ...
- **Conserves scientific** data properties alongside visualisation considerations
- No databases or dedicated servers are required, just http



And MOC for discovery

View, Access, Discovery of DataCube

DataLink

- Link a list of RESOURCEs to a DataSet via a small votable giving url, media type, semantics, description, size of the linked resource
- Useful for linking :
  - Additional metadata (provenance, fine characterisation)
  - Auxiliary or associated files
  - Other formats
  - Previews
  - Services applied to the dataset
- DataLink service descriptor

A resource to describe a service (included in main DAL service or {links} resource response)

SODA parameters

Similar to SIAV2, but with a different meaning

ID : ivo ID of the dataset to process

POS : spatial area to be matched

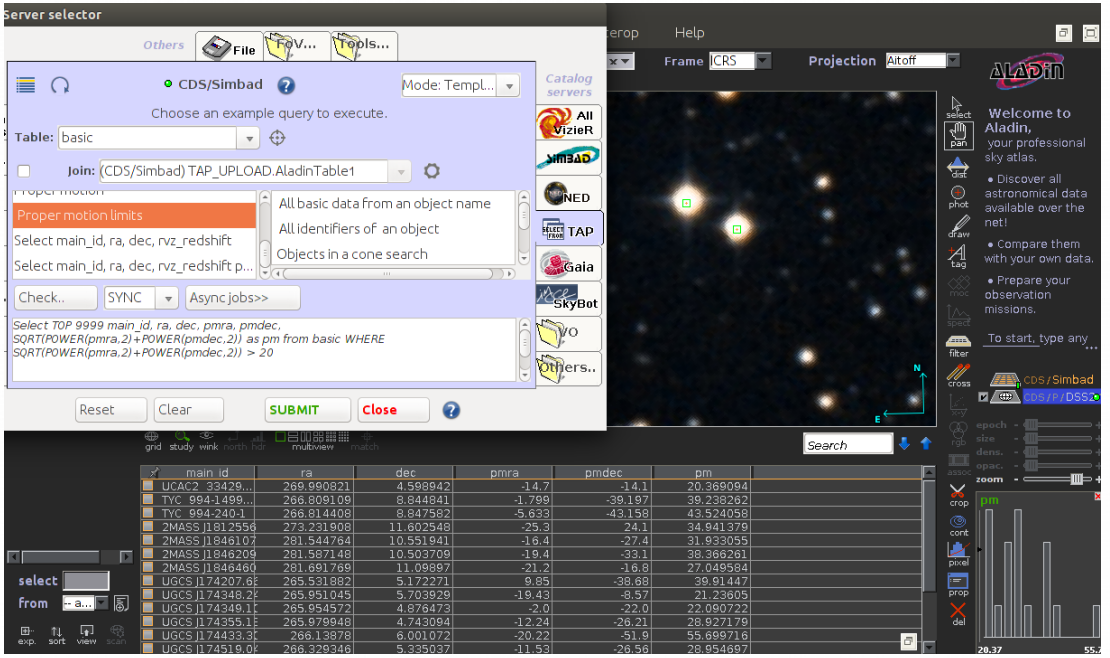
BAND : spectral range to match

TIME : time range to match

POL : selection of polarisation states

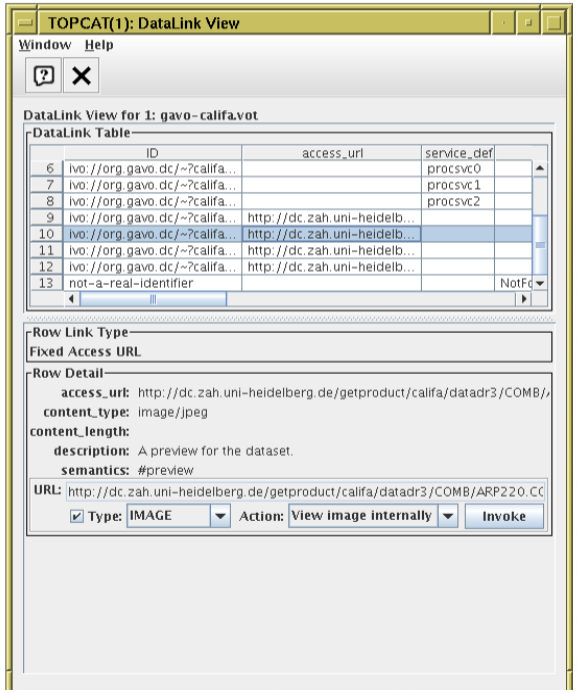
Interface : Aladin

- ASTERICS developments : discovery tree, SIA interface, DataLink interface and SODA interface (see above)
- ASTERICS development :
  - TAP interface
    - ADQL control and assistance
    - Synchronous, asynchronous
    - JOIN
    - UPLOAD



Interface : TOPCAT

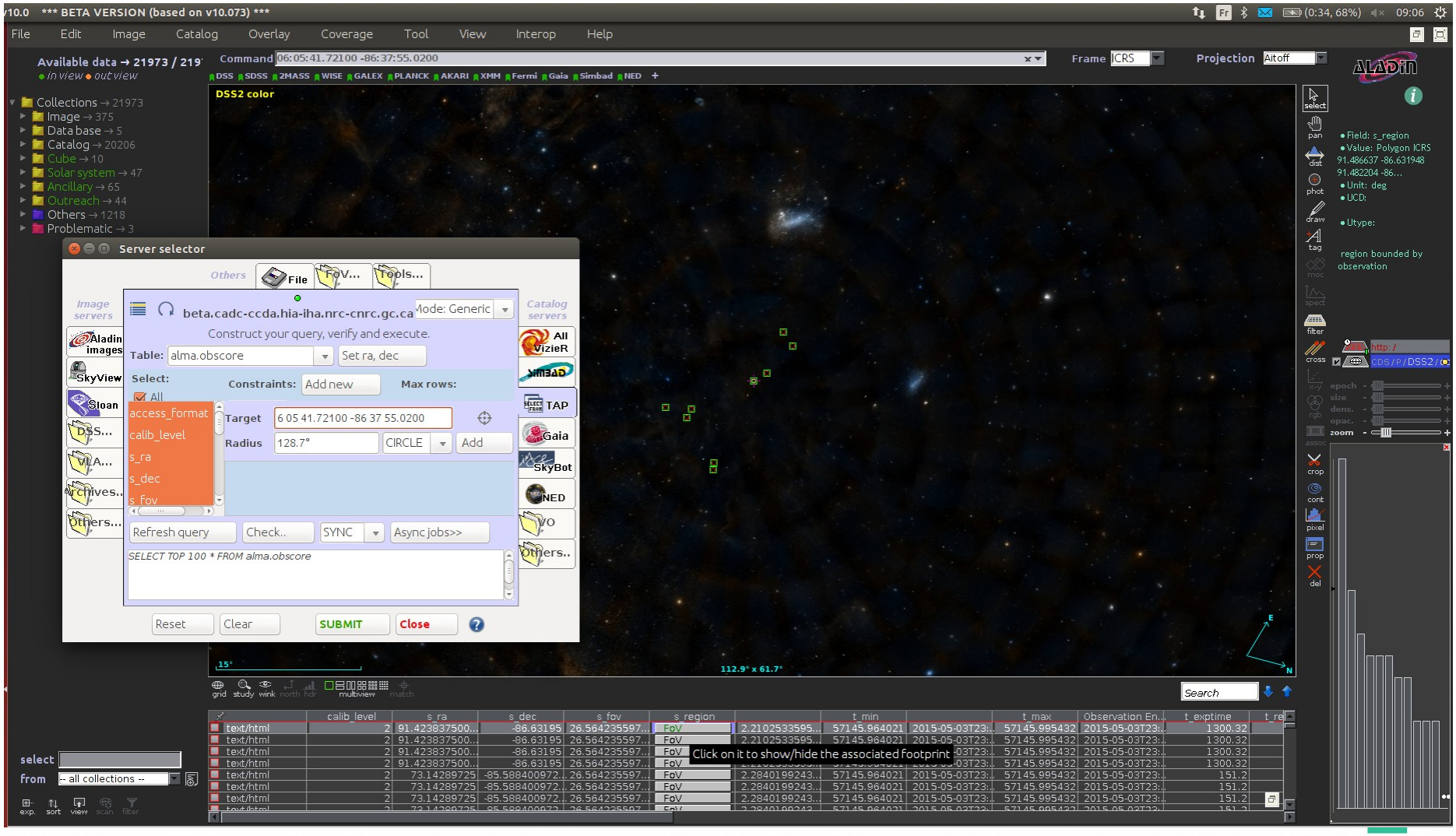
- TOPCAT is an interface to ObsTAP
  - Discovery, SAMP to other Tools
- TOPCAT takes into account DataLink
  - Feedback
  - Improvment proposals



ALMA science archive

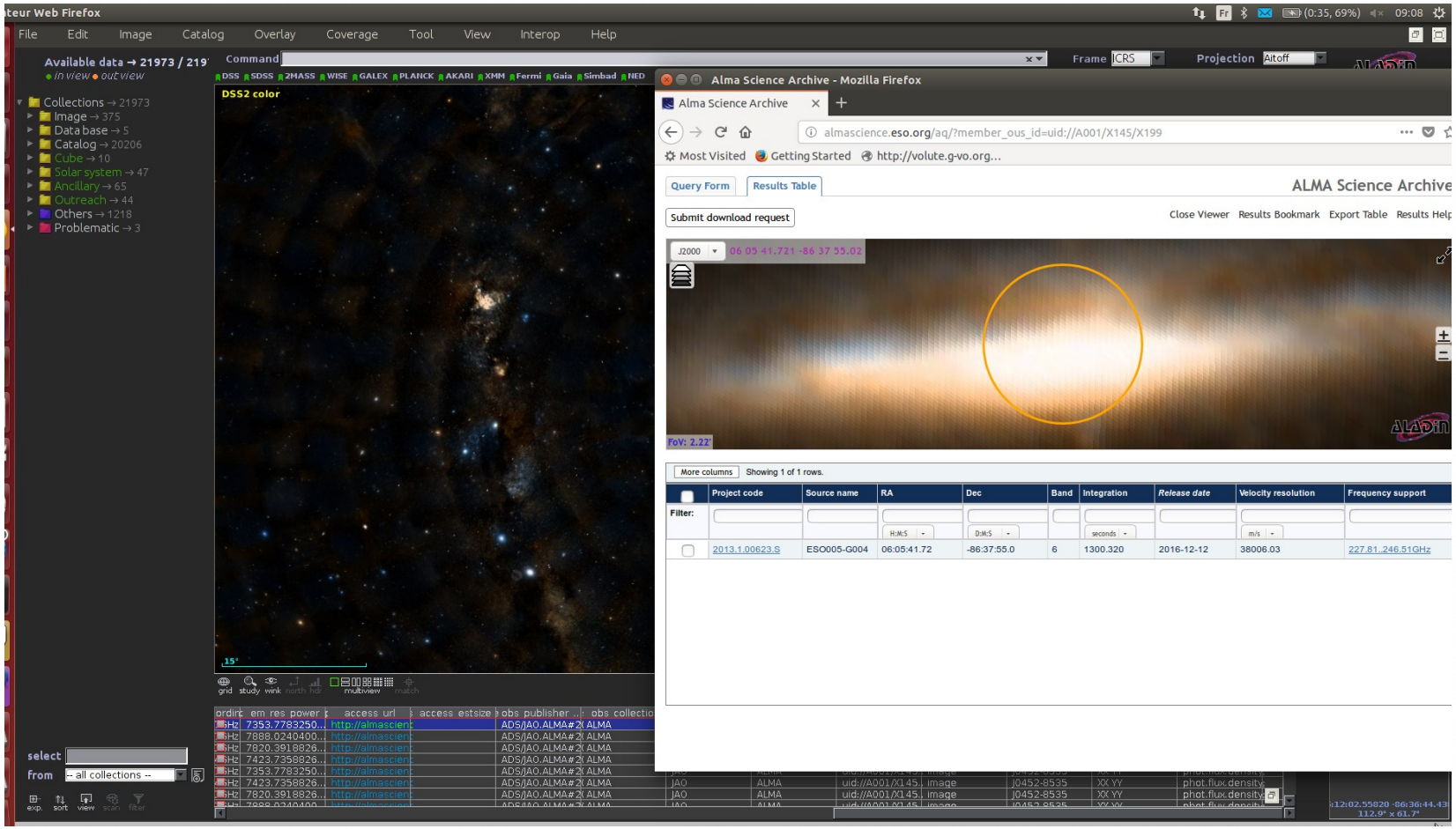
- ALMA provides interferometric data in the millimetric domain
- Science archive has a specific interface
- Imaging data are not available everywhere
- Alma science archive slowly releasing VO interface blocks :
  - An ObsTAP service is available (see next slides)
  - MOC is available
  - Integrated in ESA Sky
  - Plans to integrate : DataLink, SIAV2, SAMP in terface
  - Service for reprocessing installed as UWS (maybe SODA extension)

ALMA science archive
Tap service within Aladin



ALMA science archive

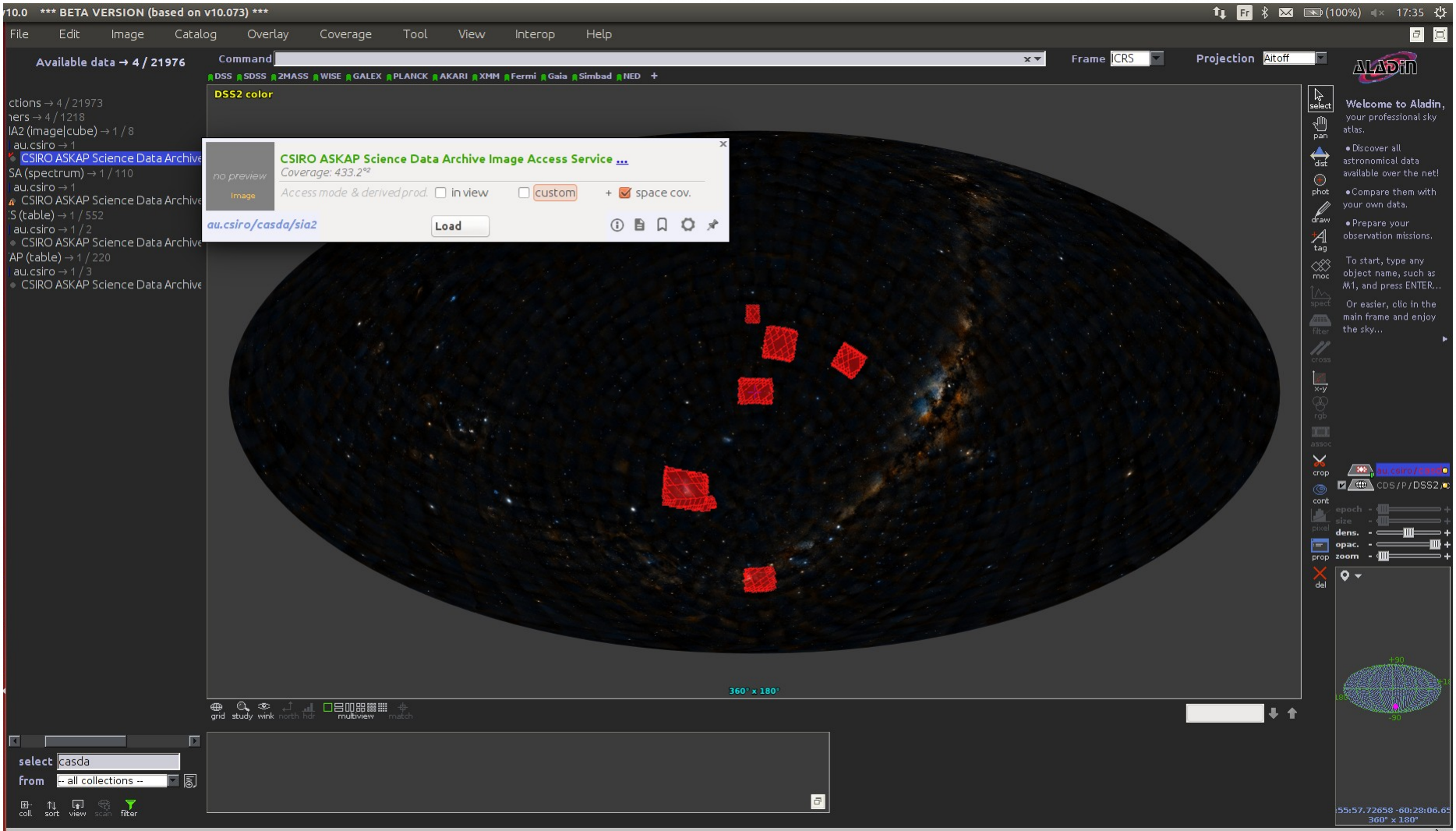
url in ObscOre response links to a page displayed in the browser
(no data download available yet)



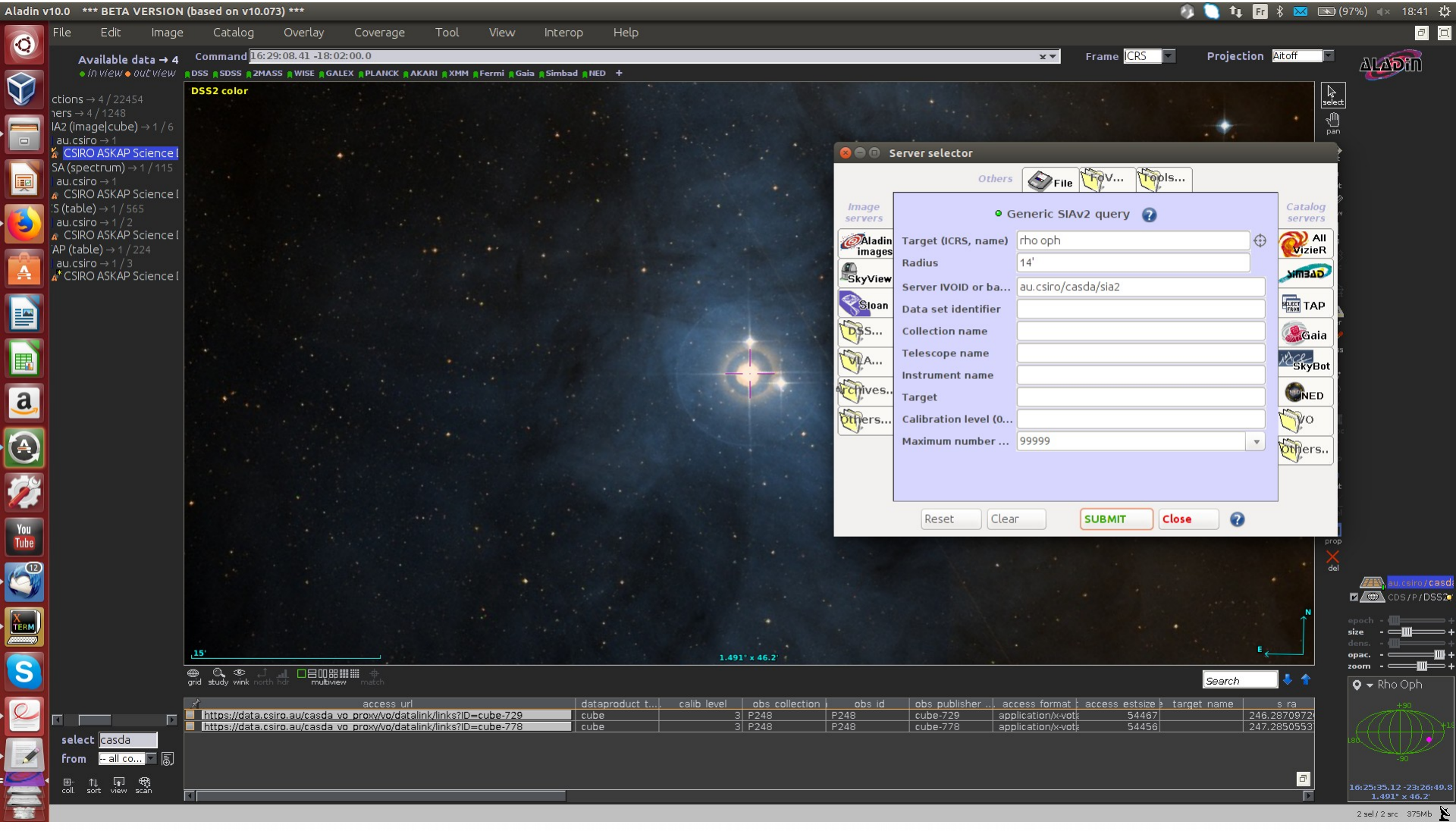
ASKAP data : CASDA services

- CASDA implemented
  - MOC
  - ObsTAP, SIA services available
  - DataLink with Cutouts
  - SODA interface
  - Scripts

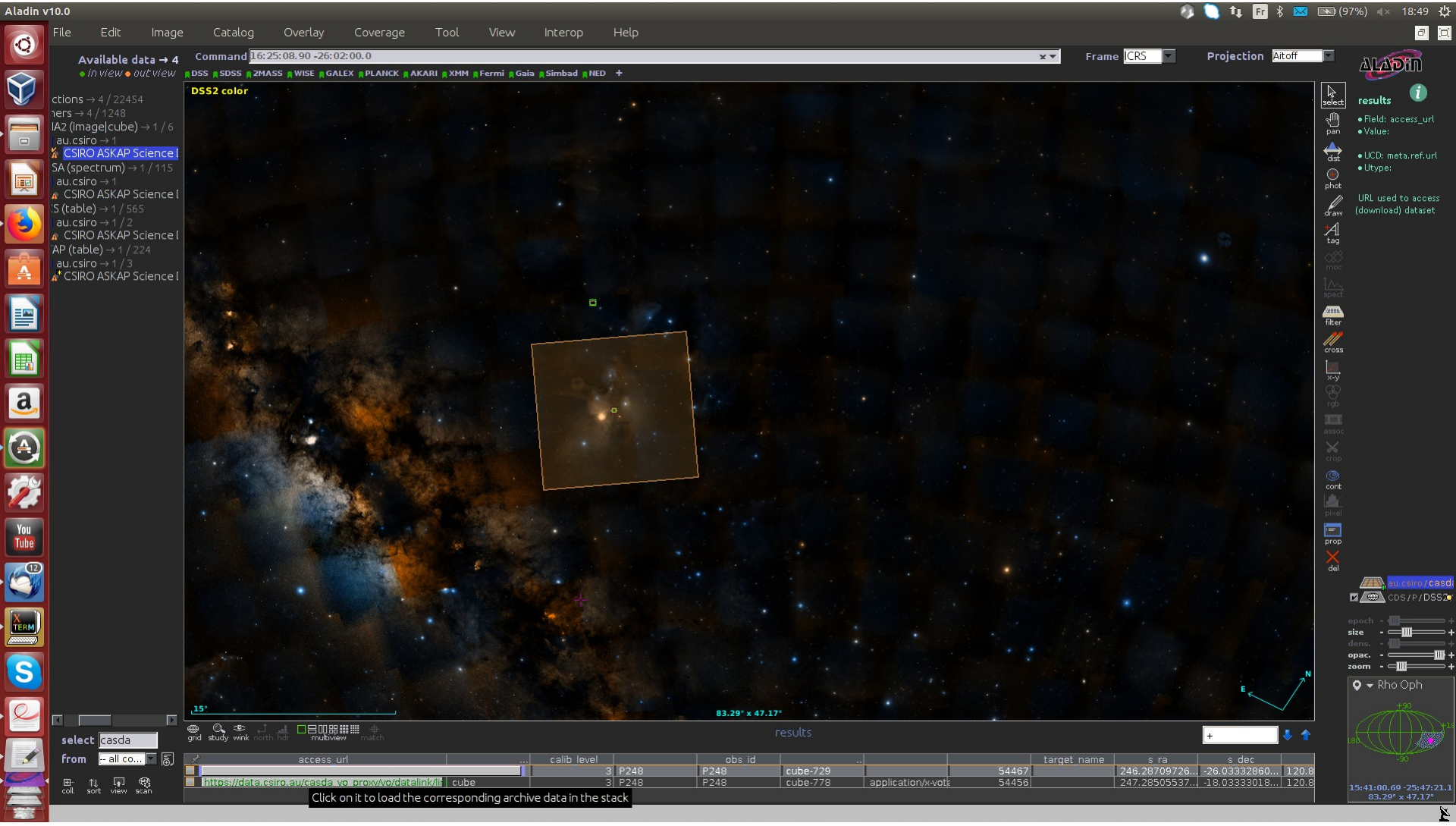
CASDA (= ASKAP data)
(allsky datacubes distribution)



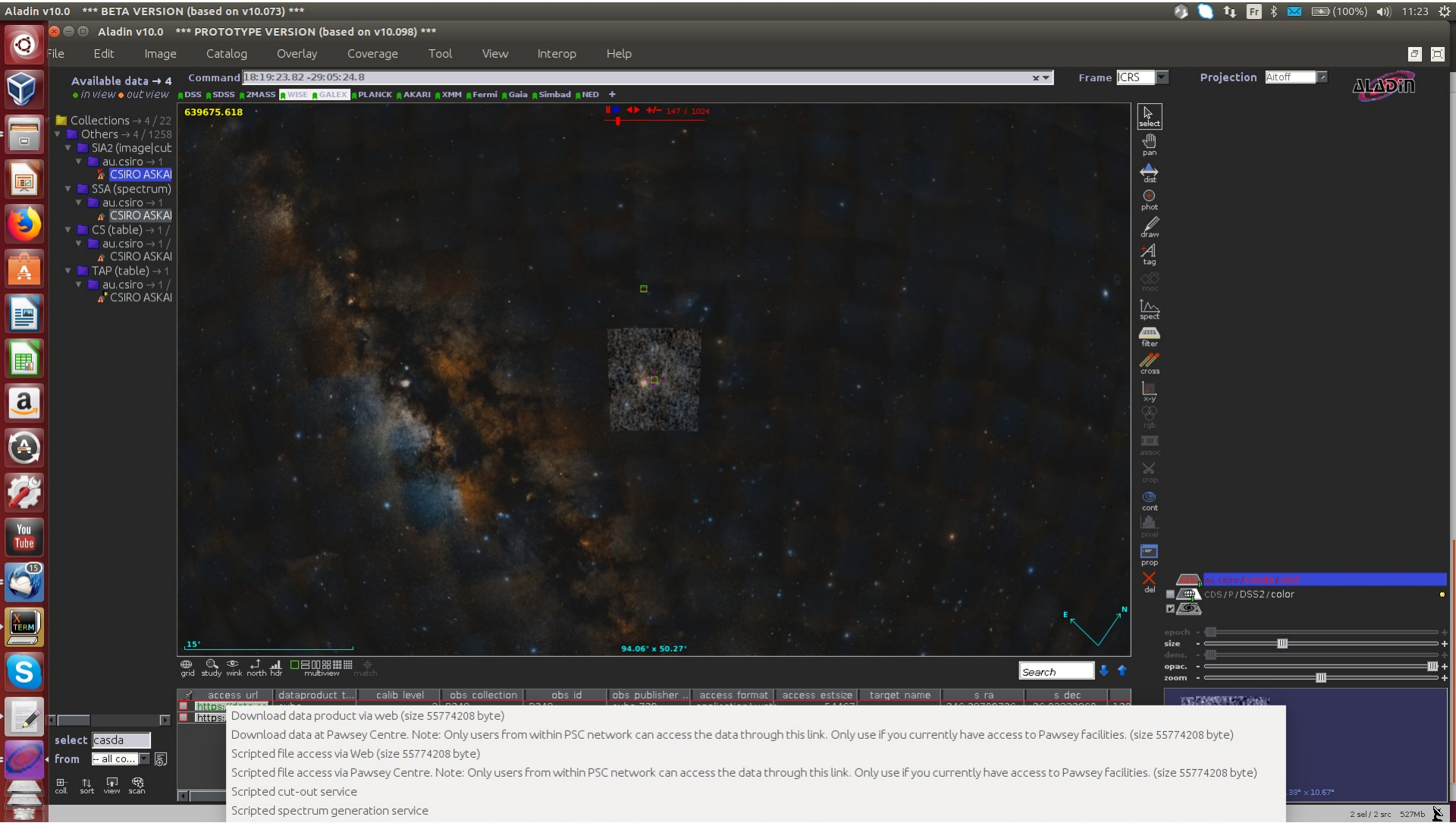
ASKAP data : CASDA services
SIAP 2.0



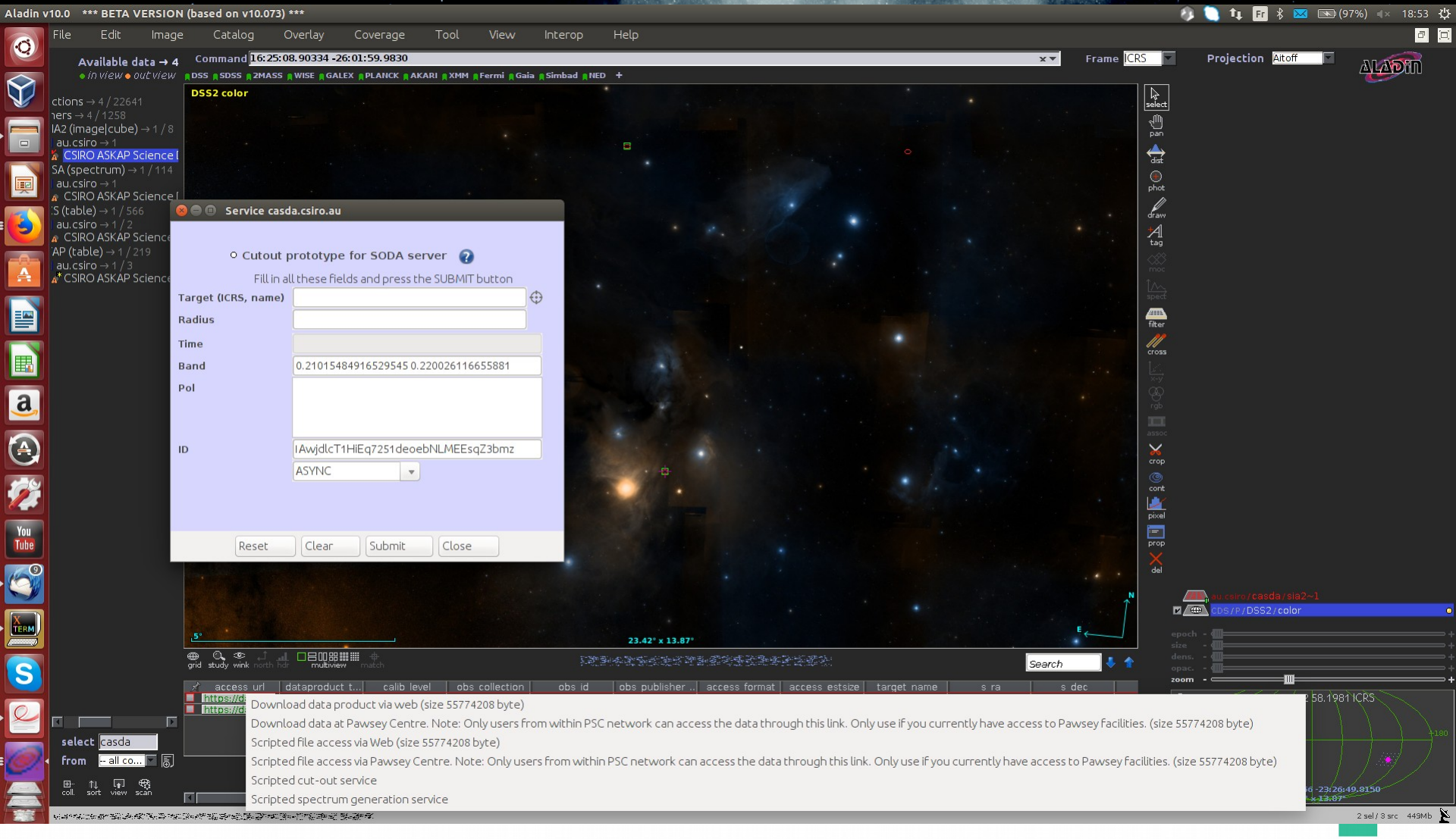
ASKAP data : CASDA services response FOV



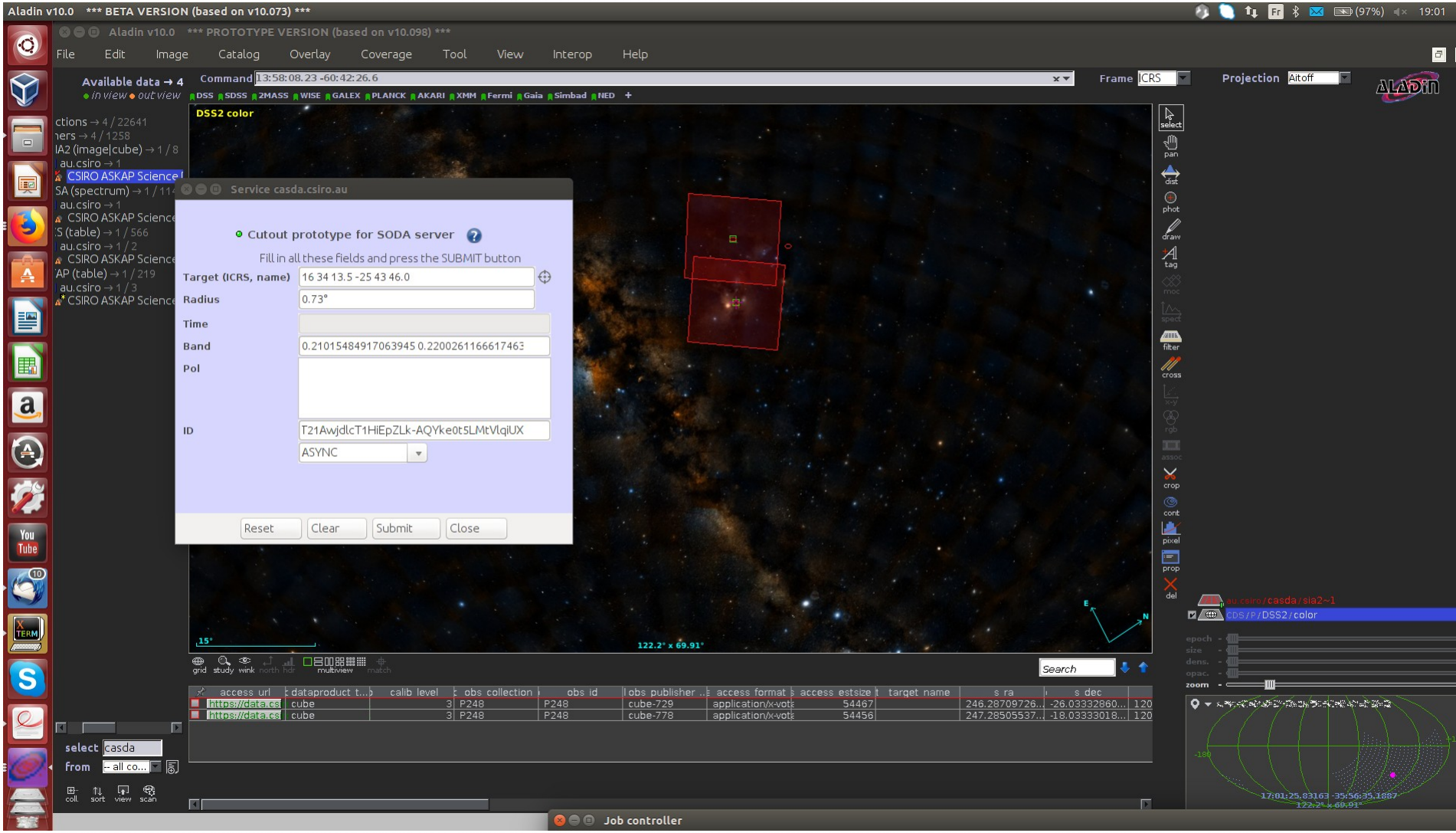
ASKAP data : CASDA services DataLink



ASKAP data : CASDA services
SODA interface

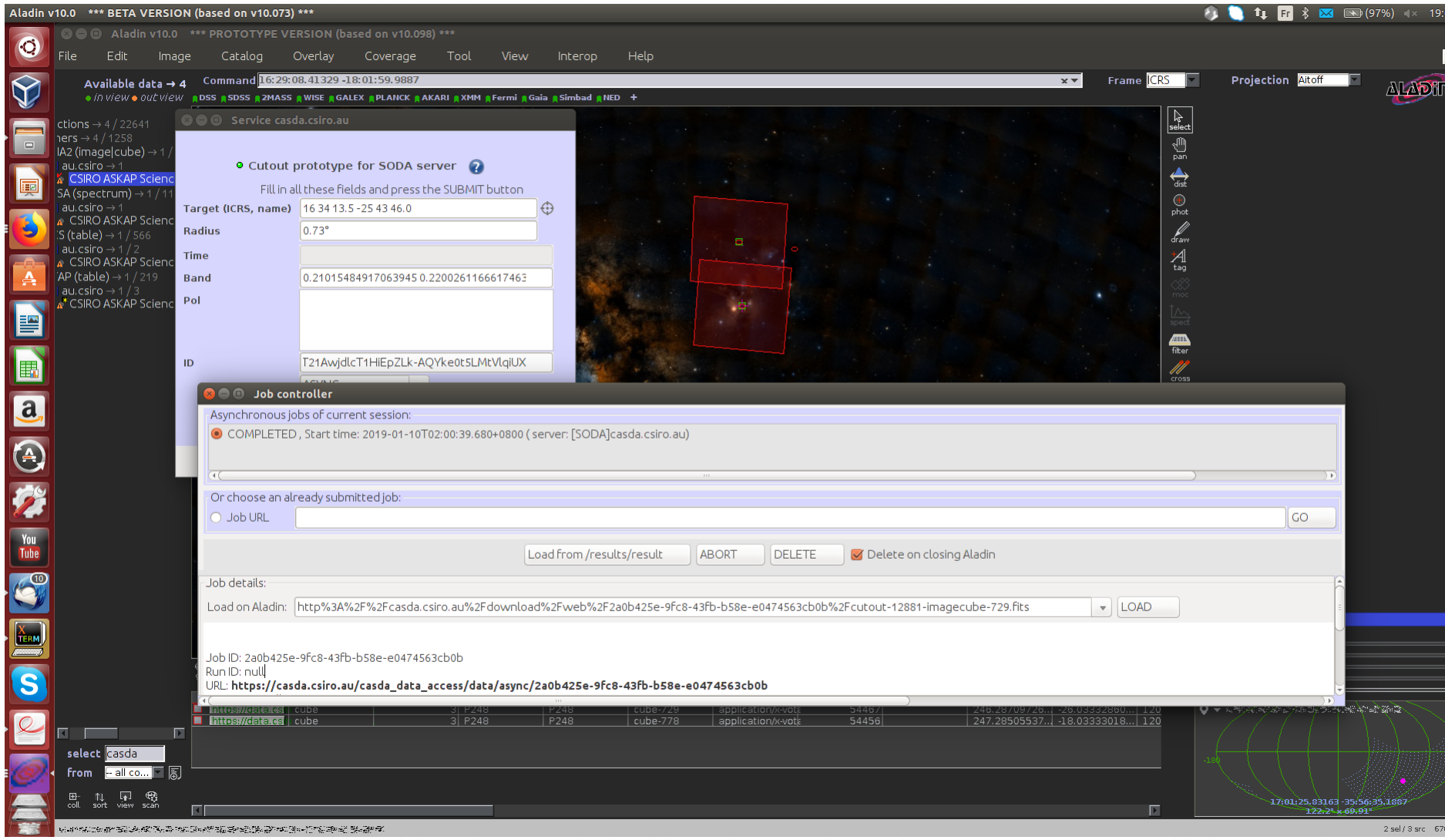


ASKAP data : CASDA services
Ready for SODA query

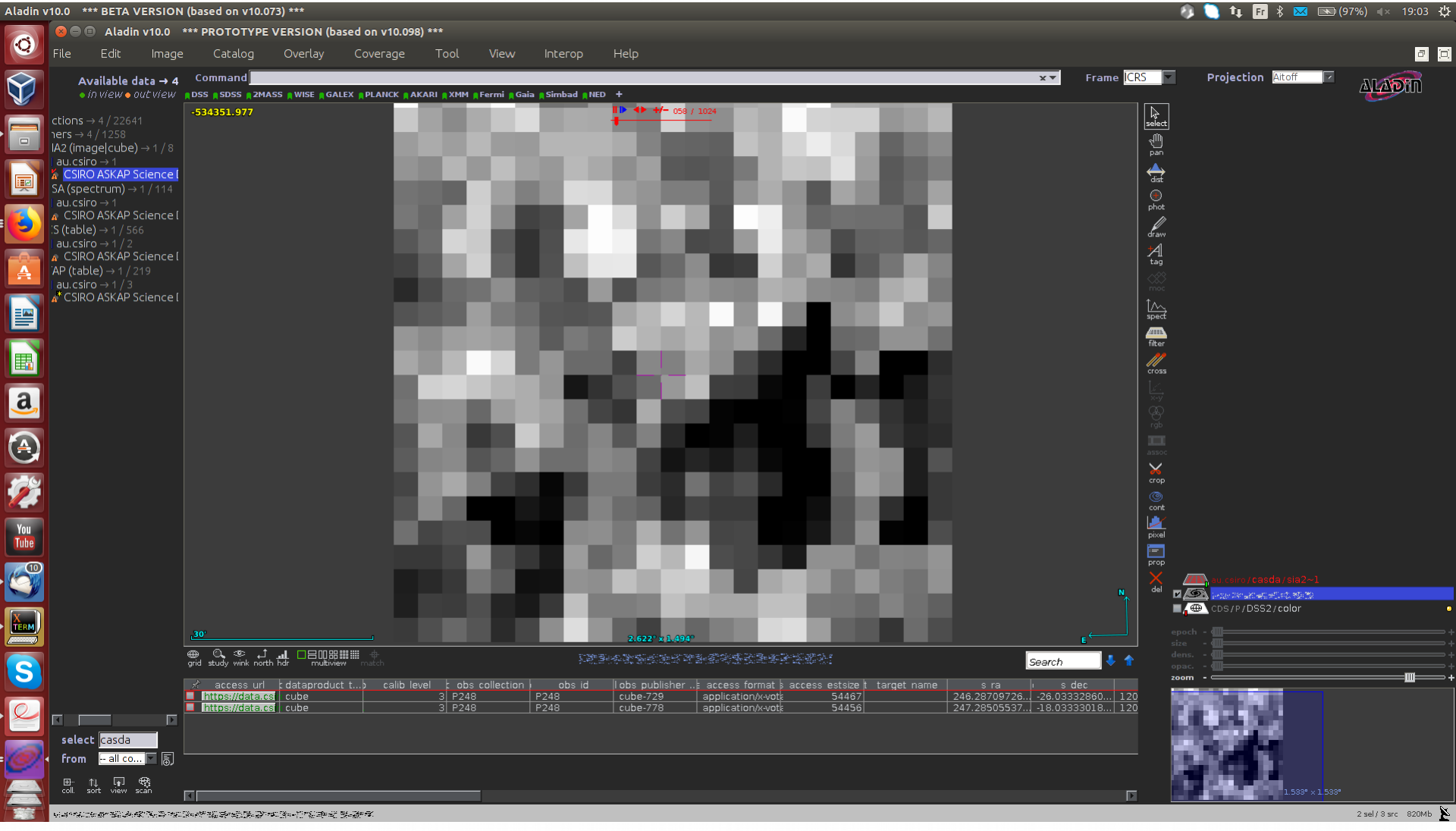


ASKAP data : CASDA services

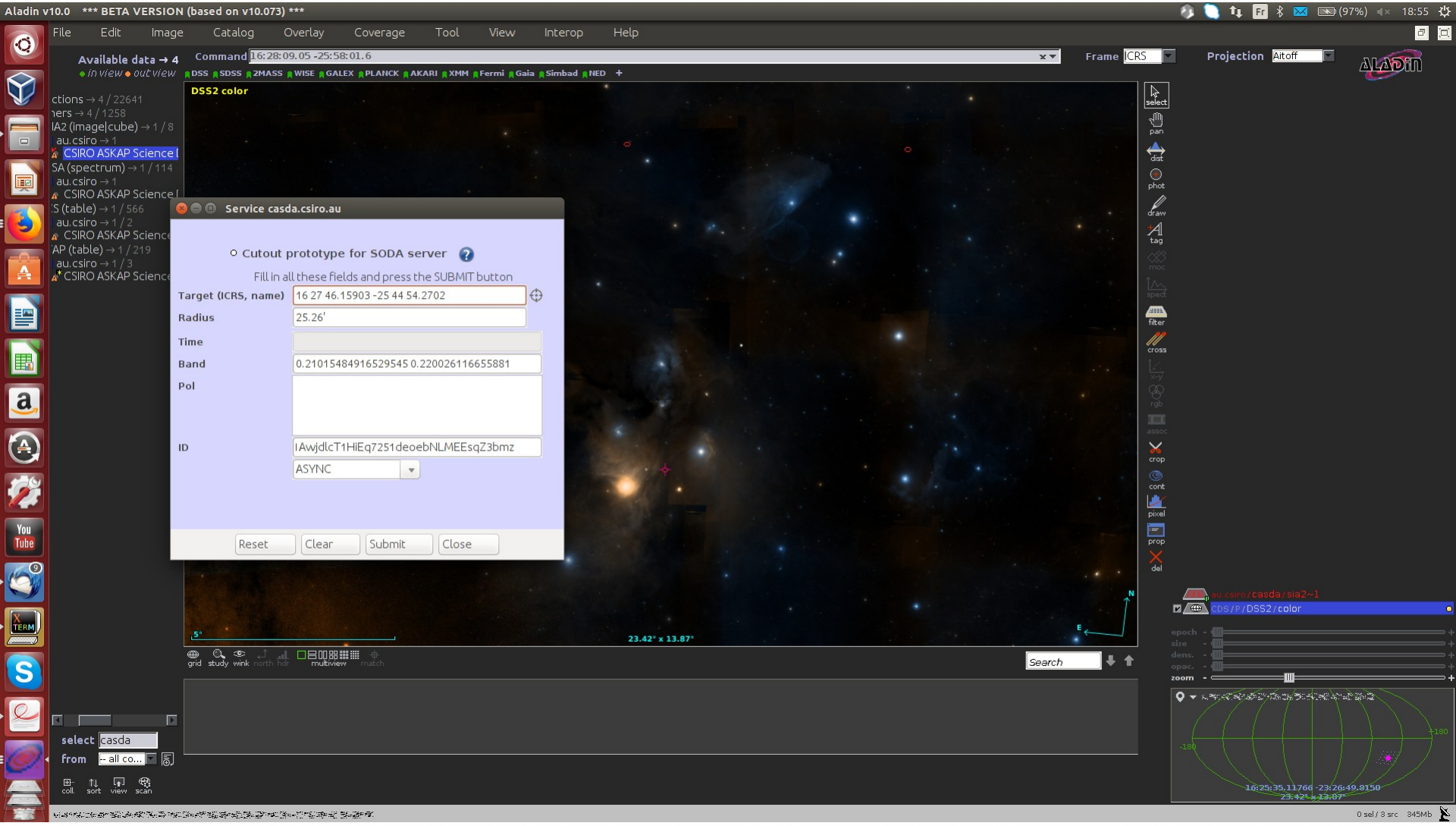
Asynchronous SODA task completed



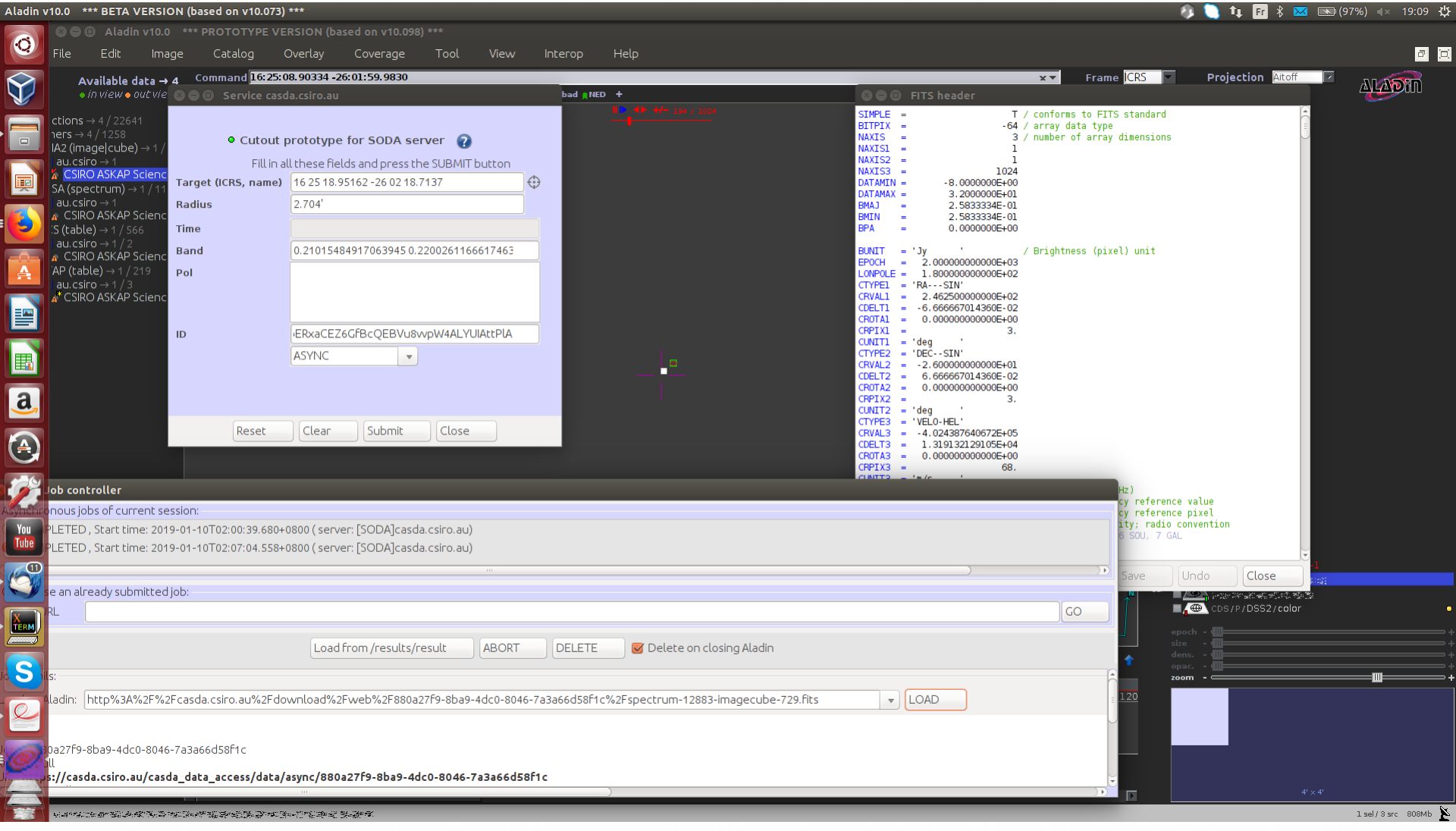
ASKAP data : CASDA services cutout cube loaded



ASKAP data : CASDA services spectrum cutout



ASKAP data : CASDA services
spectrum cutout

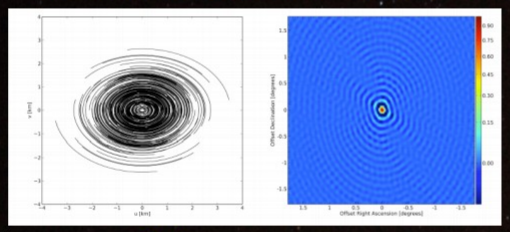


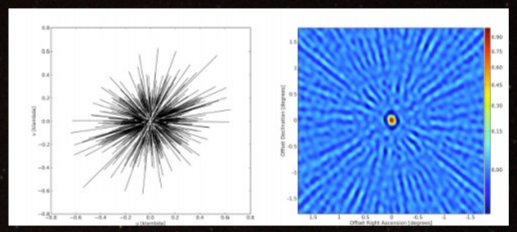
Visibility data ?
(direct discovery or via

Data Types: Visibilities

ASTRON

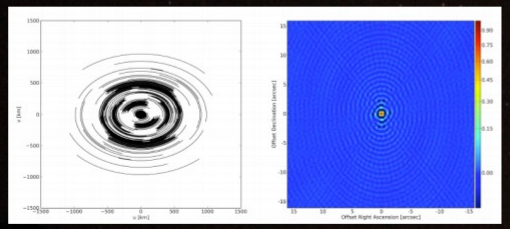
- Interferometer: Creating virtual aperture up to ~2000km (λ/D)
- Incomplete coverage
- Typical use of Radio Telescopes: imaging depending on science
- LOFAR: Projection effects (No moving dish/mirror)

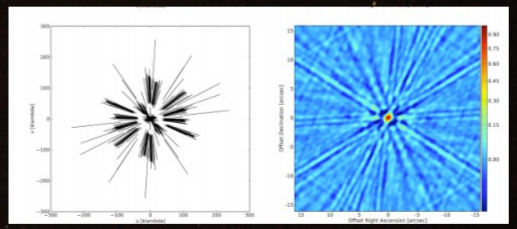




6 hours

1 second



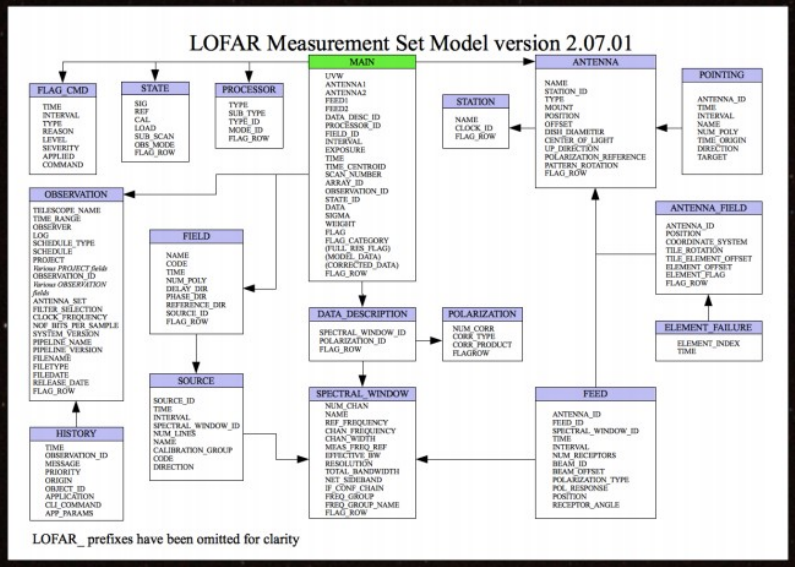


Visibility data ?
(LOFAR measurement sets model :)

Data Types: Visibilities

ASTRON

- Stored as CASA Measurement Sets
  - 10 MB - 100 GB in size, per 192 kHz frequency band
  - Usually 400-488 per observation
  - Relational structure
  - Each cell can have multiple dimensions (pol x freq)



Visibility datamodel ?:

derived from cube datamodel
inspired by TimeSeries datamodel

DataProduct
Either regular
Ndimage
Or sparsed cube

Sparse Visibility : set of uv, time
Spectral, polarization,
Photometric tuples (ND points)

Mixed visibility : set of
Spectro-polarization Ndimages
for some sparsed uv, time