Radio and millimetric Data in the Virtual Observatory : State of the art and open questions

F.Bonnarel (CDS)

On behalf of … many Interop Partners

Summary

- 0) political context
- I) science ready data / visibility data
- III) relevant VO achievments or ongoing work
  - Dicovery : obstap/Siap2
  - Access : DataLink and SODA
  - Interfaces : Aladin, TOPCAT
  - Discovery + Access + View : MOC and HiPS
- IV) working implementations
  - HiPS cubes distributed by CDS
  - ALMA implementations
  - ASKAP data : CASDA implementations
- V) What next ?
  - « extend » « science ready data » implementations
  - Discovery of visibility data
  - « fine grain » characterisation of visibility data
  - Integrating Visibility data processing in VO landscape
  - Visibility data represention ?

Radio cubes in the VO political context

- « Multi-D » IVOA priority driver
  - Set of protocols
- ALMA,LOFAR, ASKAP in VO landscape.
  - First IVOA standards implementations
  - Asterics discussions in Europe
- SKA perspective
- ESCAPE CEVO in Europe : ESFRI projects / VO experts
  - Asterics/Escape meeting in Strasbourg

Radio data in the VO

Science ready data (imaging cubes)

versus

Visibility / raw data

Radio data Cubes (science ready)

- ALMA
- LOFAR, ASKAP, JIVE

  → Towards SKA
- various surveys at CADC or CDS
- Two different strategies :
  - 2 steps strategy

    → data discovery by selecting criteria on metadata

    →Data acces : extraction of « useful » data in discovered « datasets »
  - AllSky discovery, access and visualisation : HiPS

Data discovery : ObsTAP

- Data discovery by selecting criteria on description attribute (metadata)
- ObsCore 1.1 (addition to ObsCore 1.0):
  - Size in pixels
  - Polarisation
- Serverved by TAP → ObsTAP
- ObsTAP : allows to select datasets on criteria such as dataproduct_type= cube and characterisation of axes
- REC may 2017

Data discovery : SIAP 2.0

- SIAP 2 : allows queries for image and cubes with PARAMETERS such as : DPTYPE = cube
- POS = « STC shape», BAND = 0.0005 0.0006, TIME = 52618 53700 etc.
- REC december 2015

Towards access :
→ DataLink

- Full data retrieval via access_reference column in ObsCore
- Or access to DataLink …. an SODA cutouts

DataLink

- {links} endpoint :

  link a list of RESOURCEs (metadata, other formats, associated data, services...) to a DataSet via a small VOTable giving url, media type, sémantics, etc..

- DataLink « service descriptor »

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

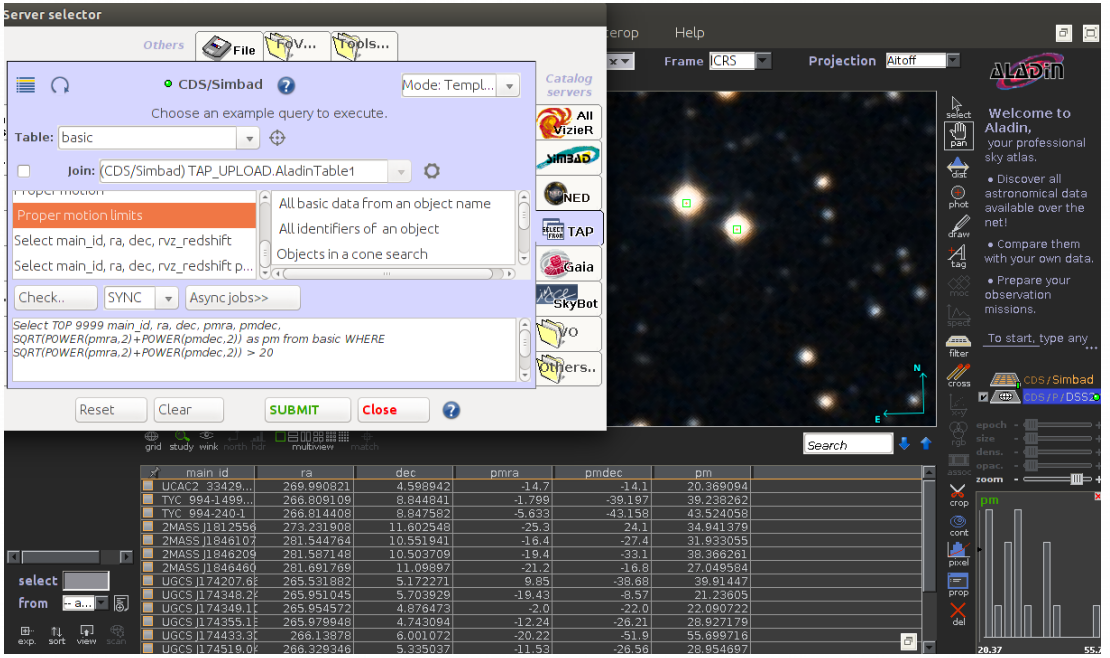
- REC June 2015
- DataLink 1.1 proposal (july 2019)

Data access

- Data access : extracting useful data from the datasets
  - Ful retrieval cumbersome (radio cubes are « huge »
  - Cutouts : SODA 1.0 REC may 2017
  - Eventually regridding, reprocessing -> SODA -next (virtual data)
  - More metadata/modelling for finer description ?

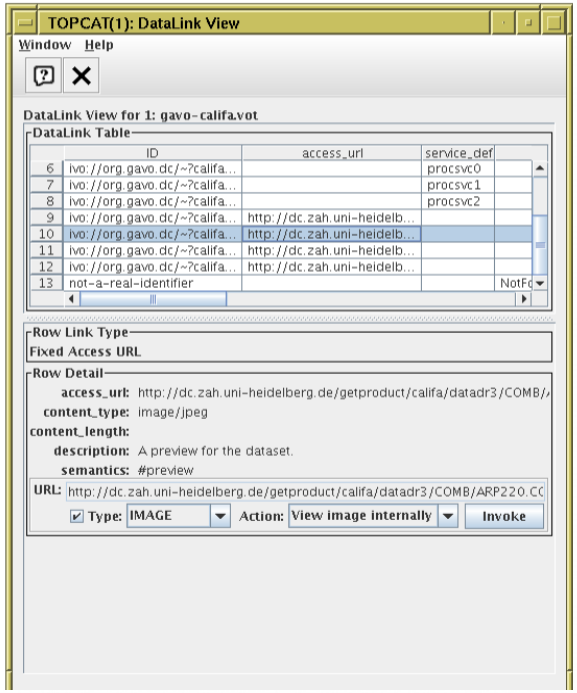
Interface : Aladin

- Recent developments : discovery tree, SIA interface, DataLink and SODA interface
- TAP interface
  - ADQL control and assistance
  - Synchronous, asynchronous modes
  - JOINs
  - UPLOAD



Interface : TOPCAT

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



VO data 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 data models
2) Other existing datamodels

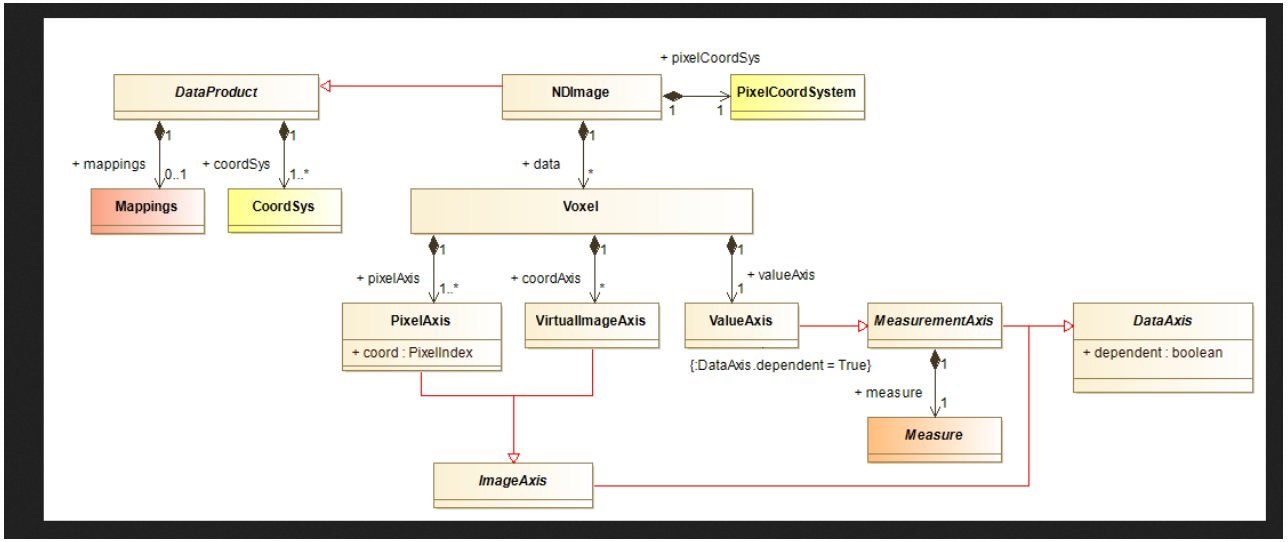
- STC : useful concepts, currently in major revision

  -> Coords, Measurements and Transformations – see below
- PhotDM : mainly for optical domain
- Characterisation 1.0 : generic model reused by other models (Obscore, Cube) . In revision
- Spectral Line Data Model : From observations to Physics
- Simulation data model and SimDAL protocol.

  → Observation/Simulation comparison

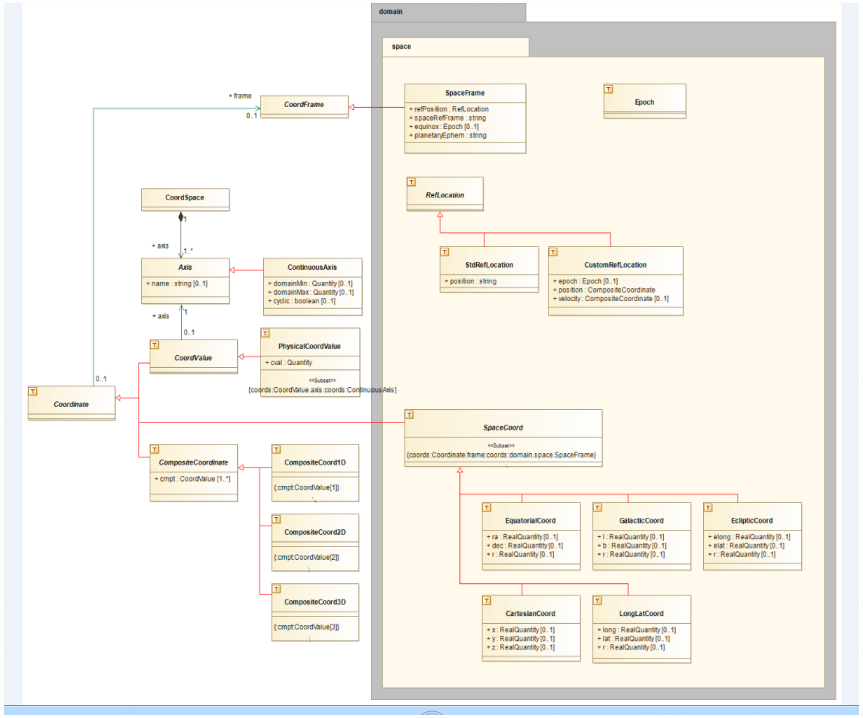
VO data modelling

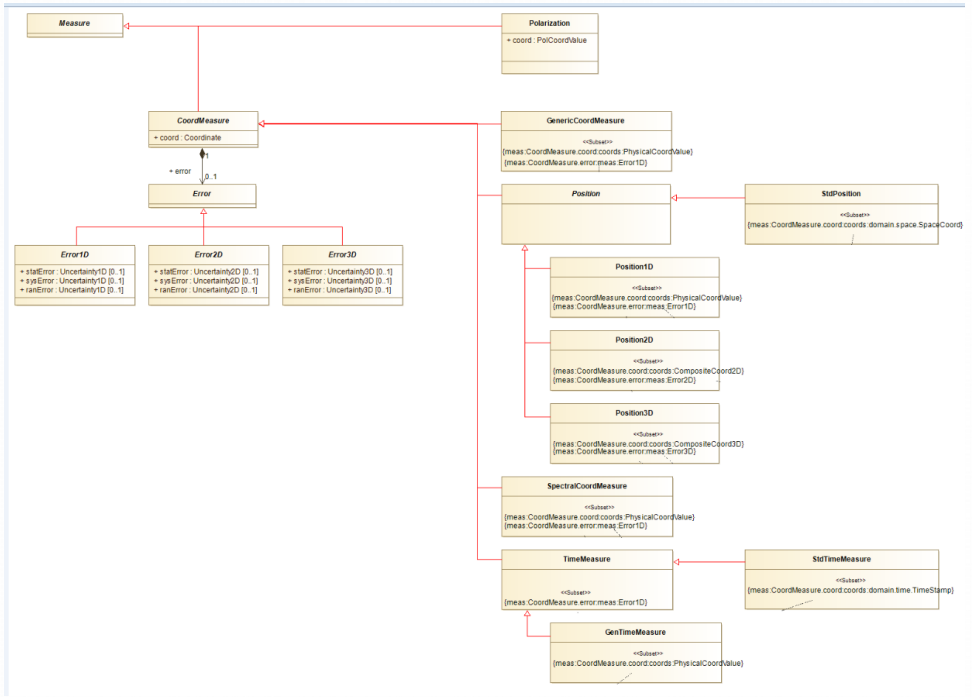
3) Cube datamodel (in progress) : ND Image



VO data modelling

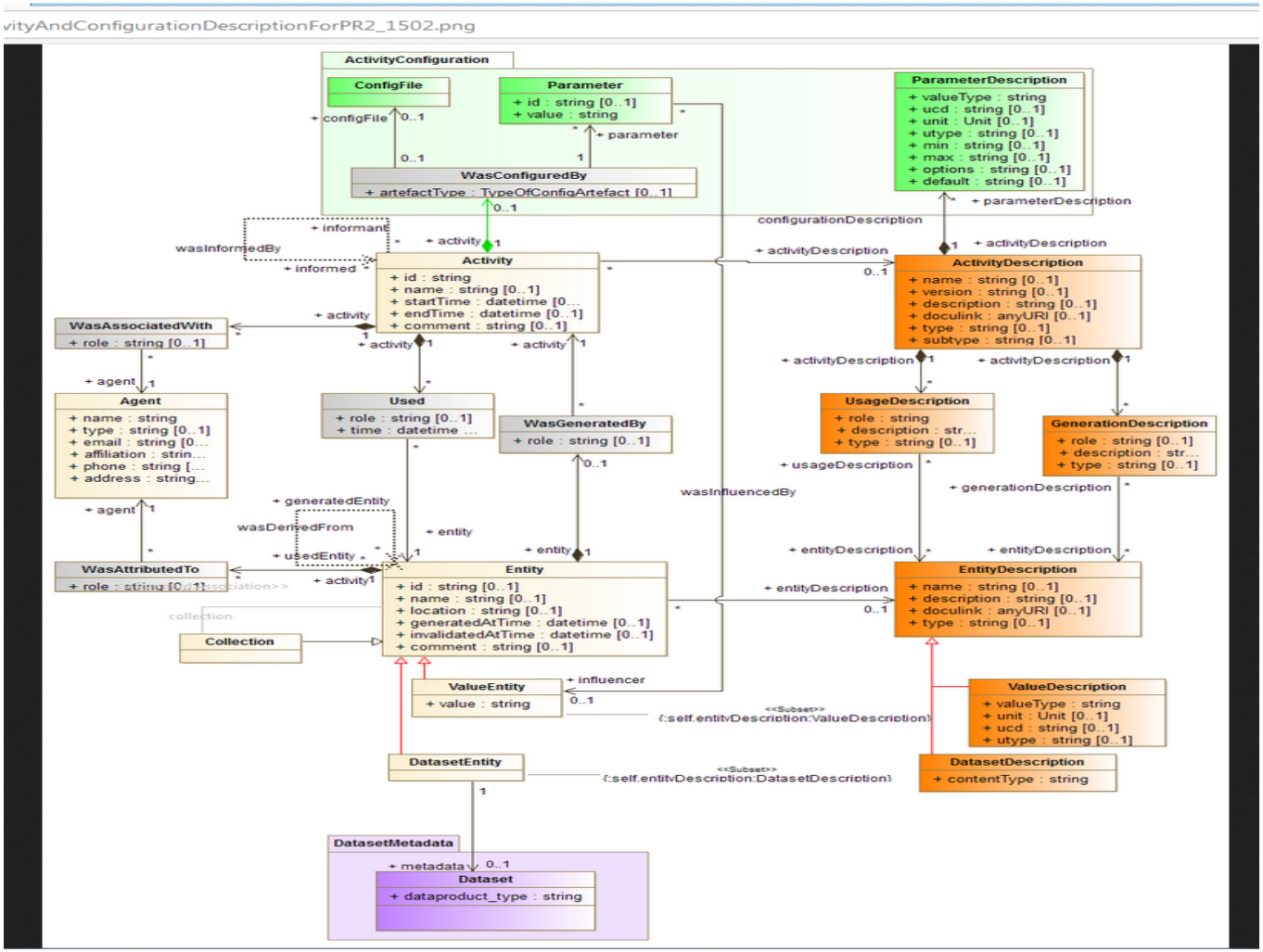
4) Coords and measurements (work in progress)





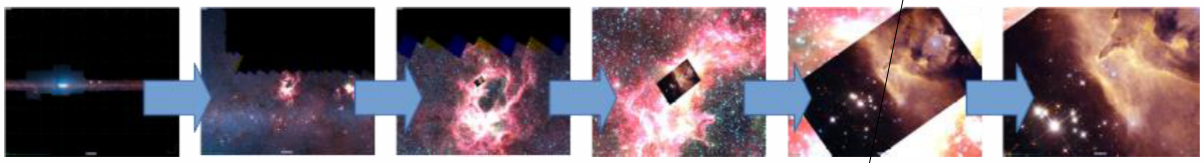
VO data modelling

5) Provenance (work in progress)



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

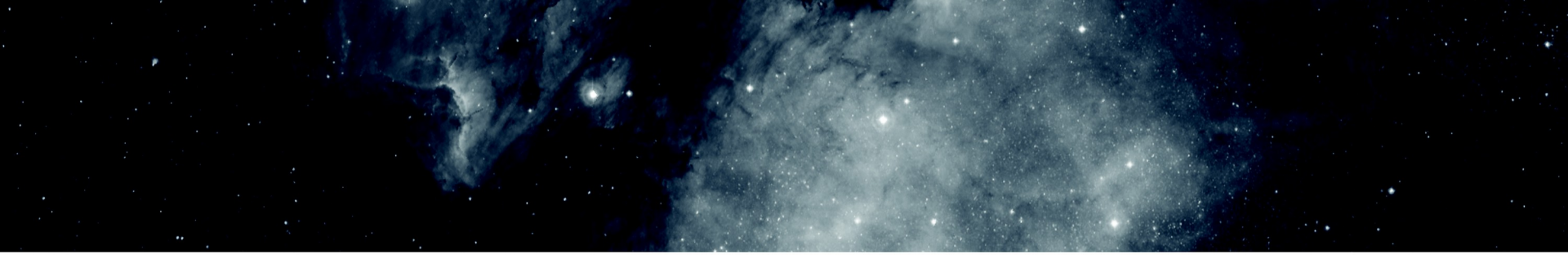


And MOC for discovery :
indexation by HEALPIX cell

View, Access, Discovery of DataCube

Don’t forget

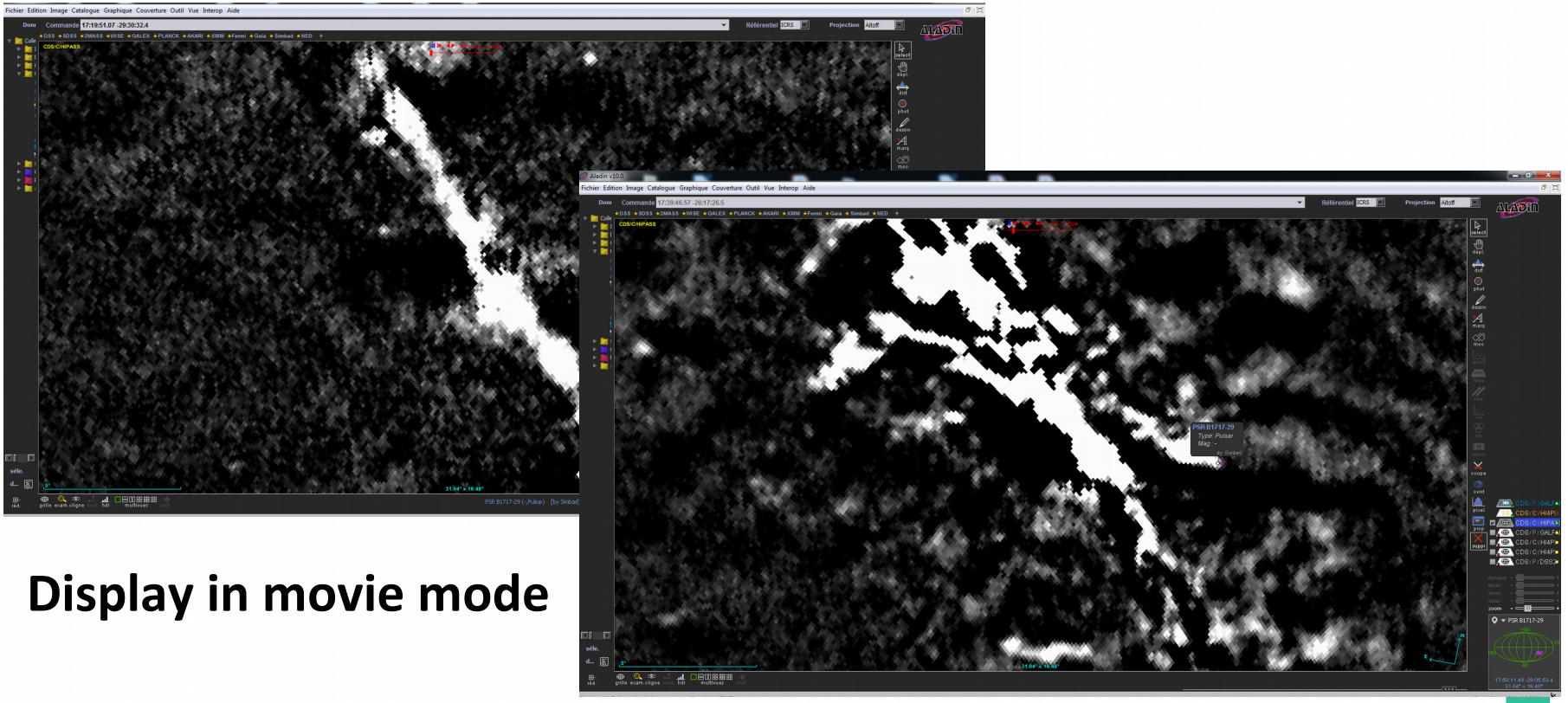
- SAMP for Desktop applications
- UWS : standard interface to manage VO Web Services
- VOSpace : for data storage
- Registry



Existing « science ready » radio data VO services

HiPS Cubes

. HI4PI (Effelsberg), GALFAHI (Arecibo), and HIPASS (Australia)

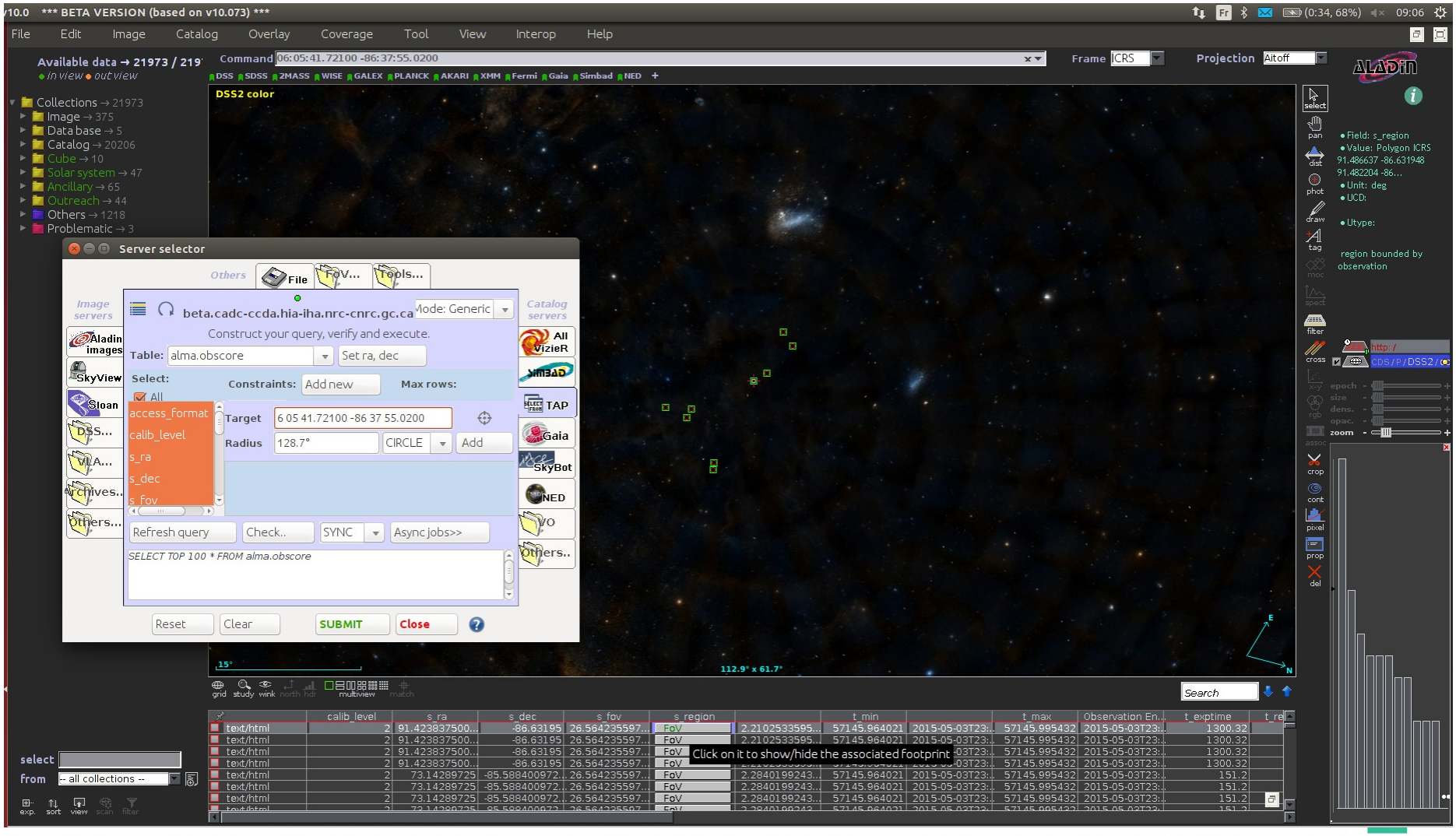


Display in movie mode

ALMA science archive

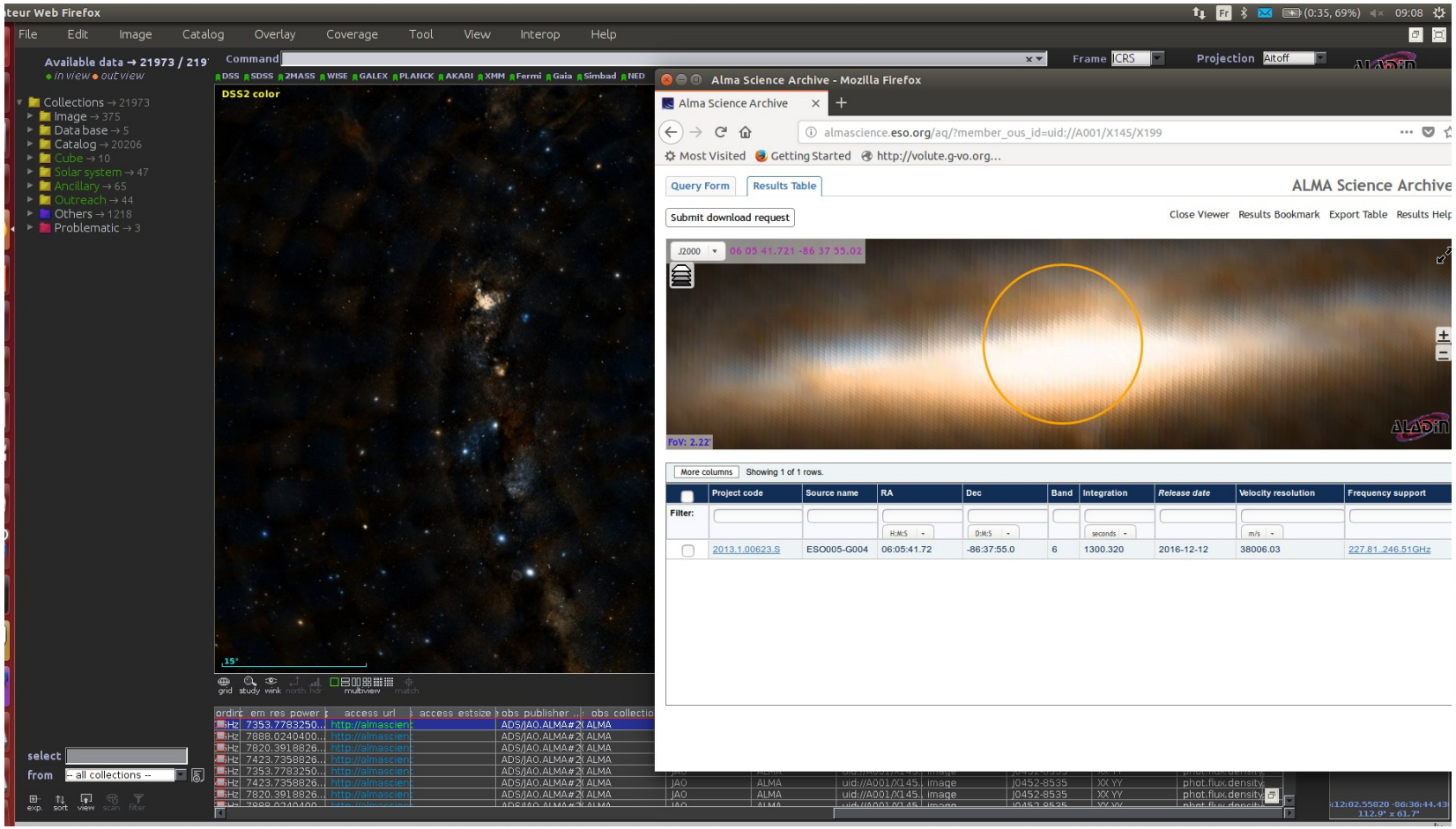
- ALMA provides interferometric data in millimetric domain
- Native interface available
- Imaging cubes not yet available for all observations
  (but in progress)
- ALMA archive delivers step by step VO interface elements:
  - ObsTAP service available (next slides)
  - SAMP in client
  - MOC available
  - Integrated in ESA Sky (with AladinLite)
  - Projects : DataLink, SIAV2
  - UWS web service for reprocessing planned

ALMA science archive
ObsTAP service in Aladin



ALMA science archive

url in the ObsCore response points to a native interface page in the browser

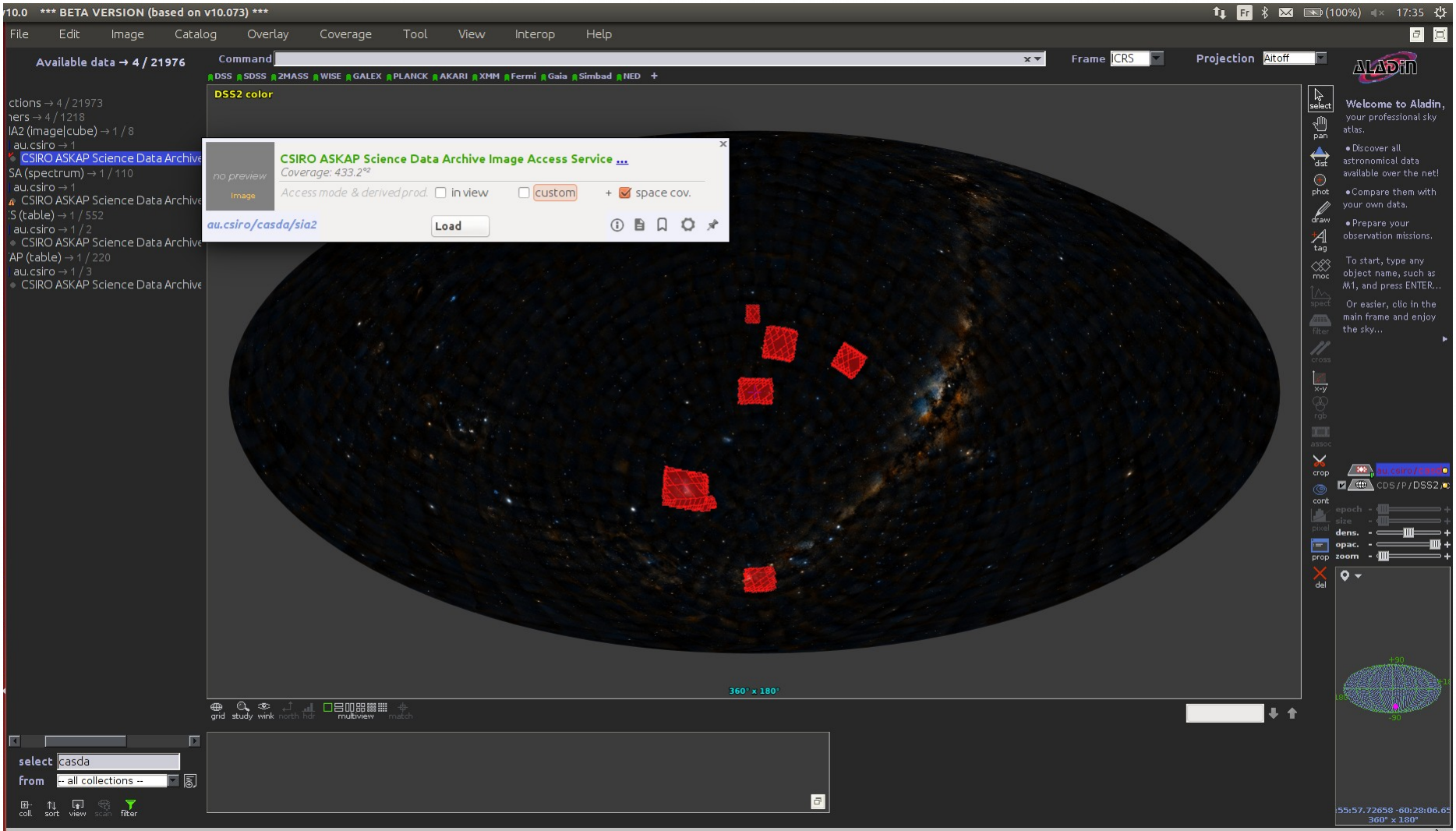


ASKAP data : CASDA services

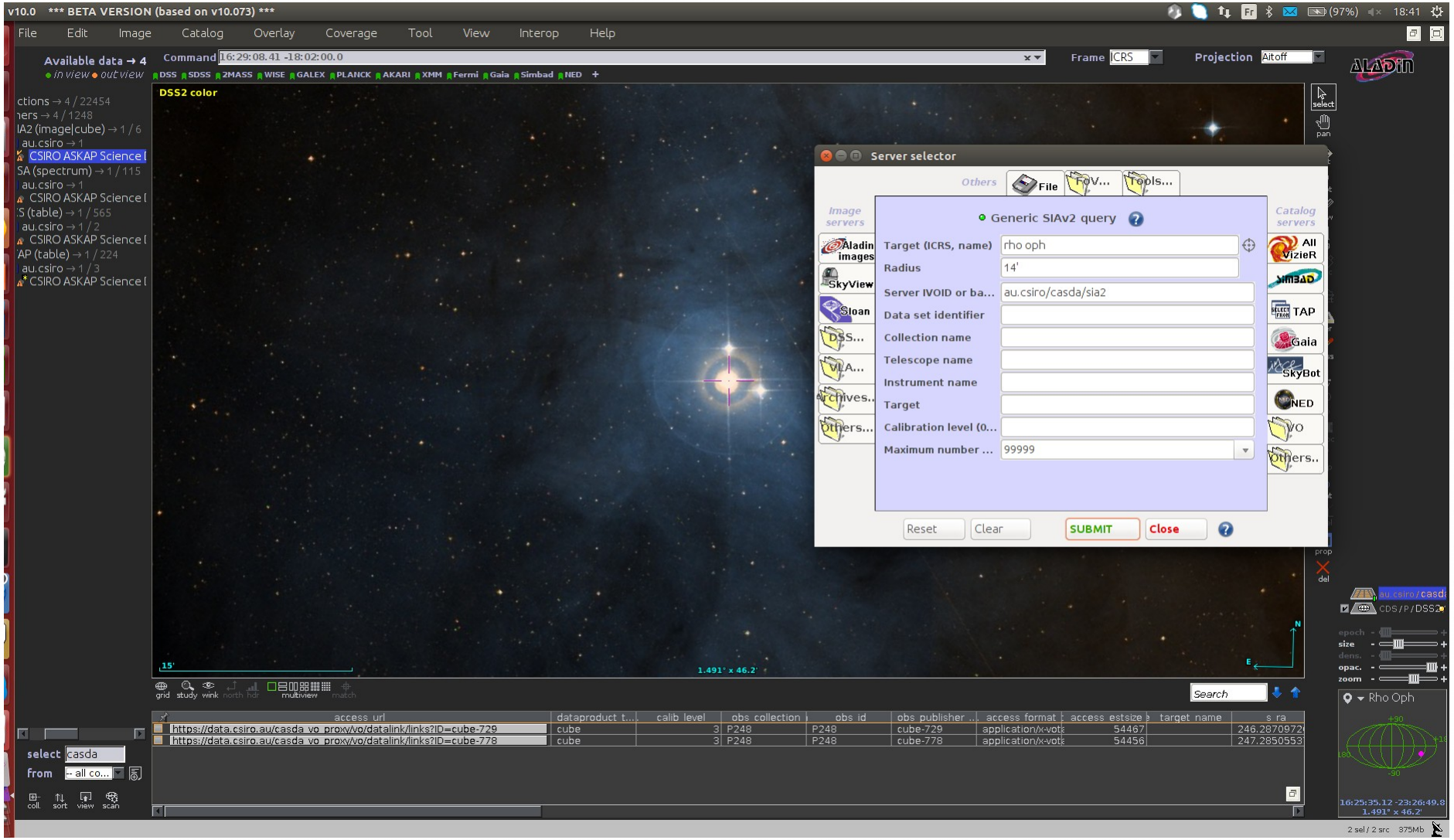
- CASDA implemented
  - MOC
  - ObsTAP, and SIA services available
  - DataLink with Cutouts
  - SODA interface
  - VO Scripts available

CASDA (= ASKAP data)

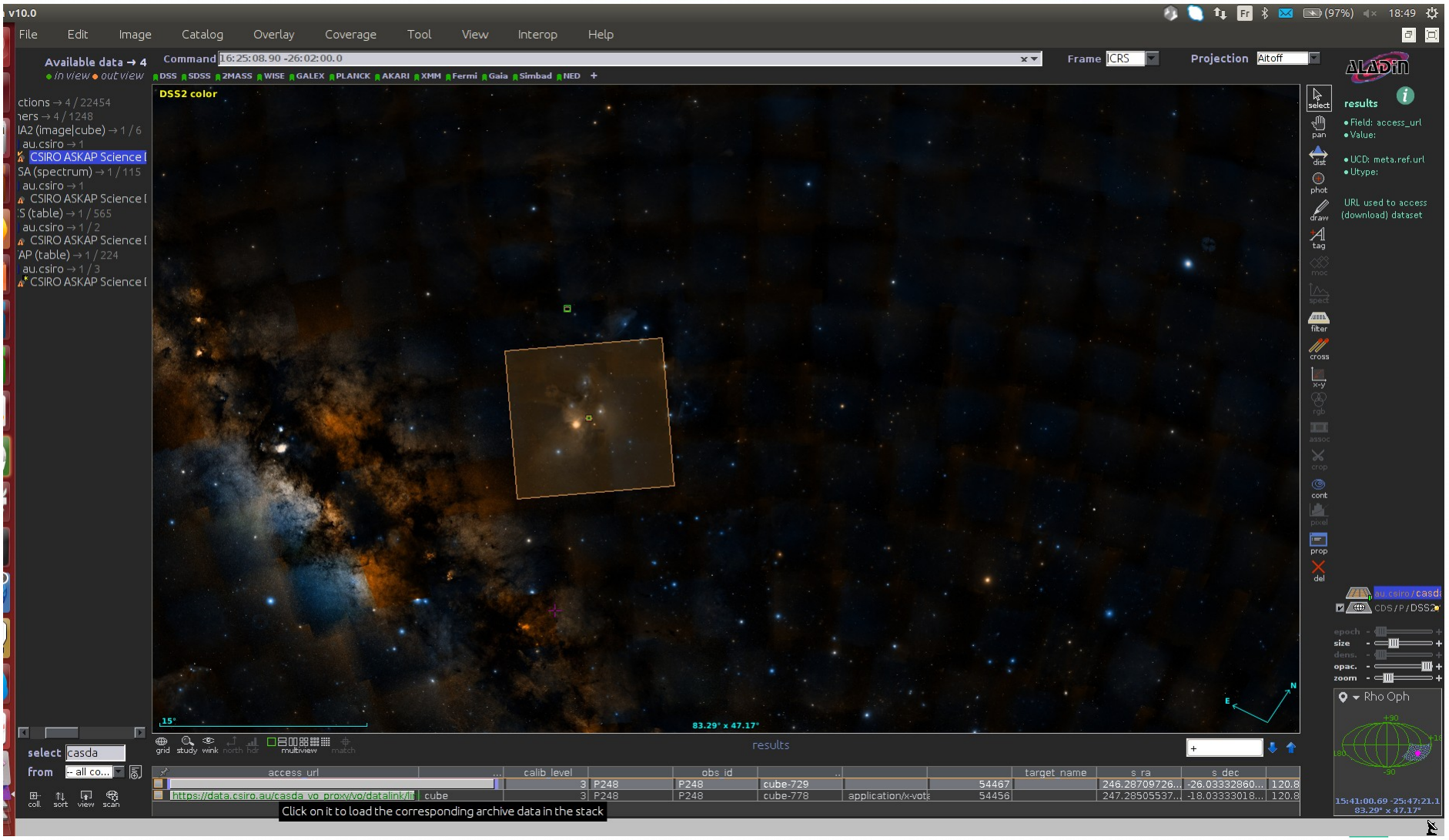
(availability of cube data on the sky)



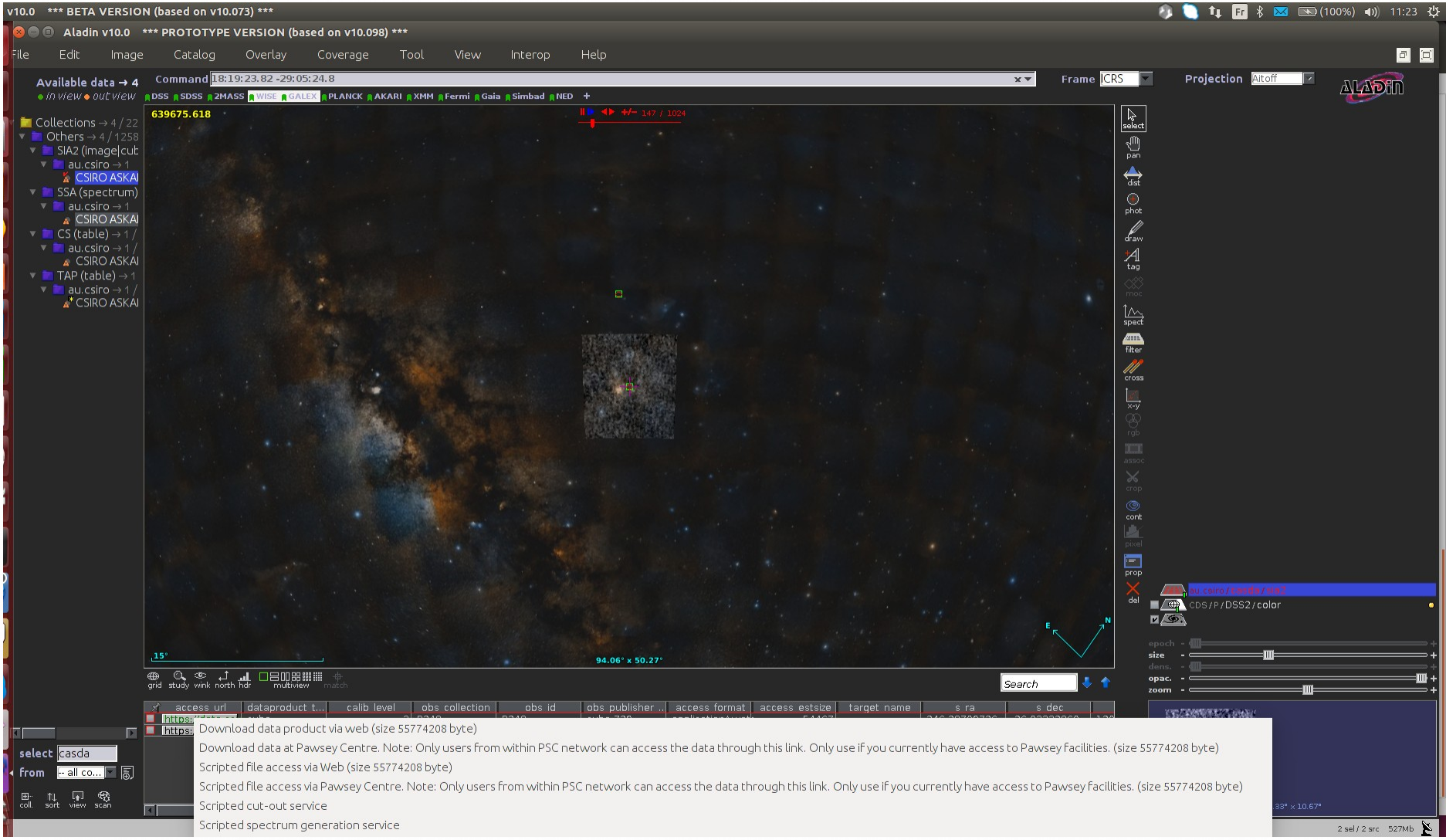
ASKAP data : CASDA services
SIAP 2.0



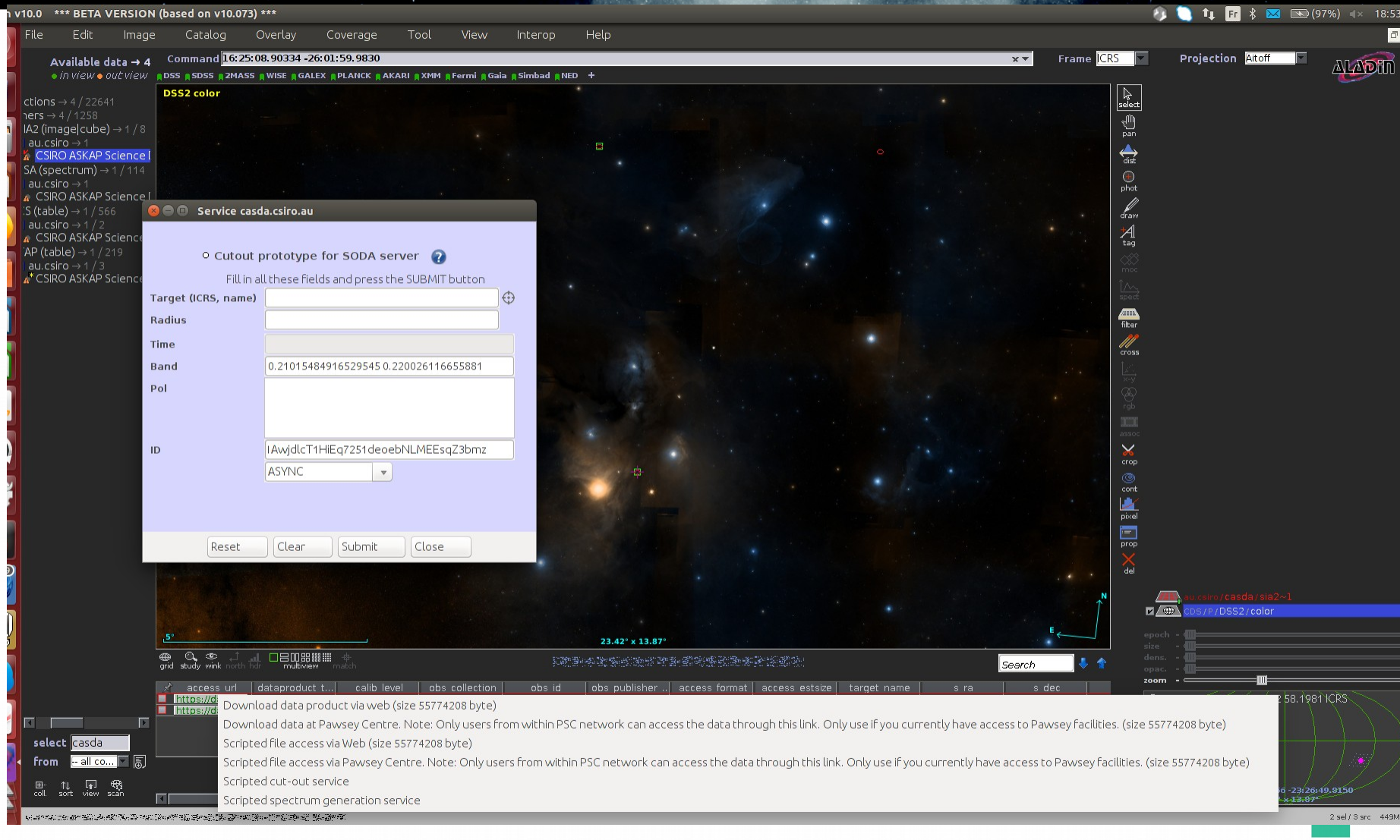
ASKAP data : CASDA services dataset field of view



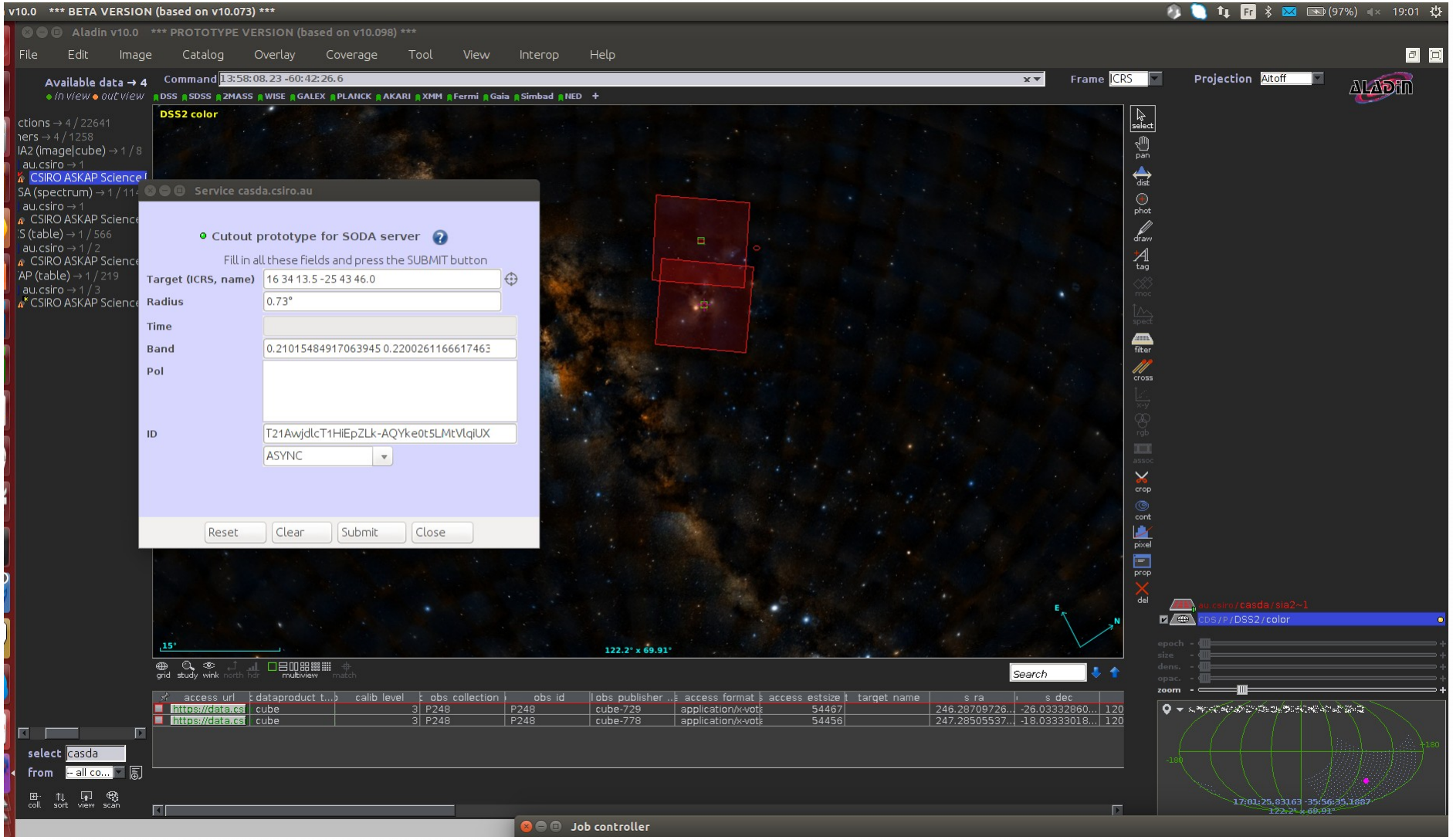
ASKAP data : CASDA services DataLink



ASKAP data : CASDA services SODA interface

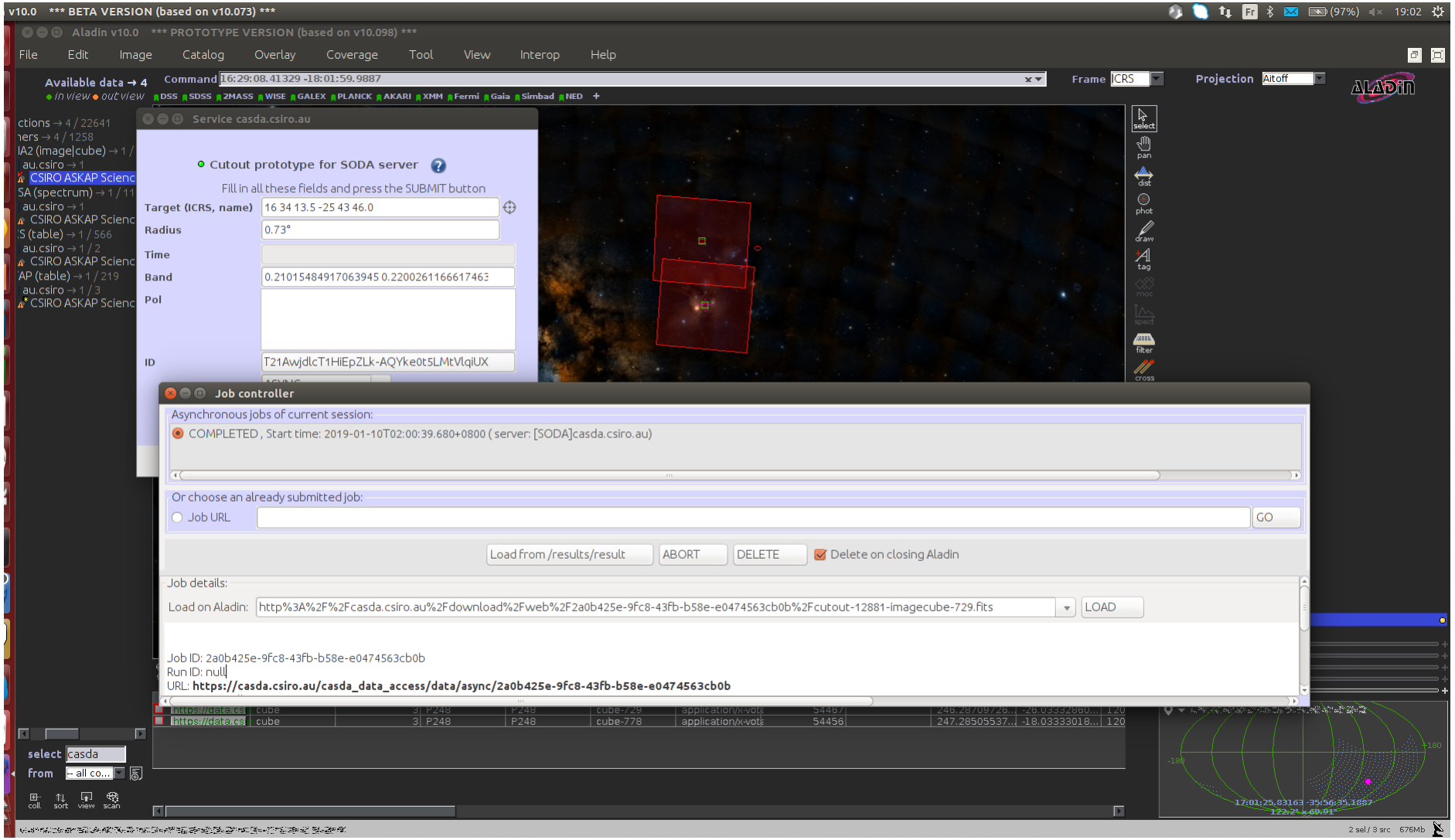


ASKAP data : CASDA services
SODA : ready for query

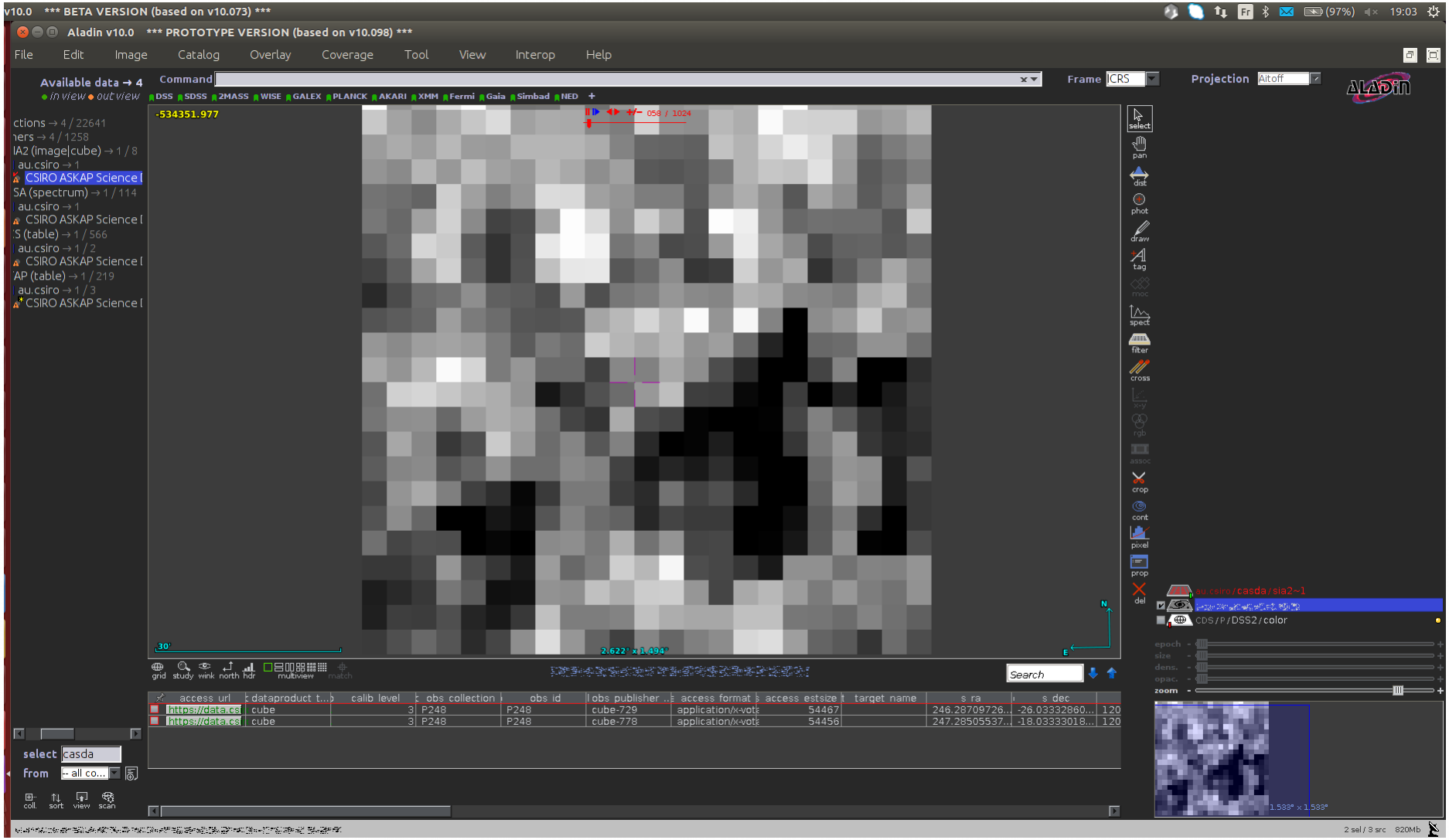


ASKAP data : CASDA services

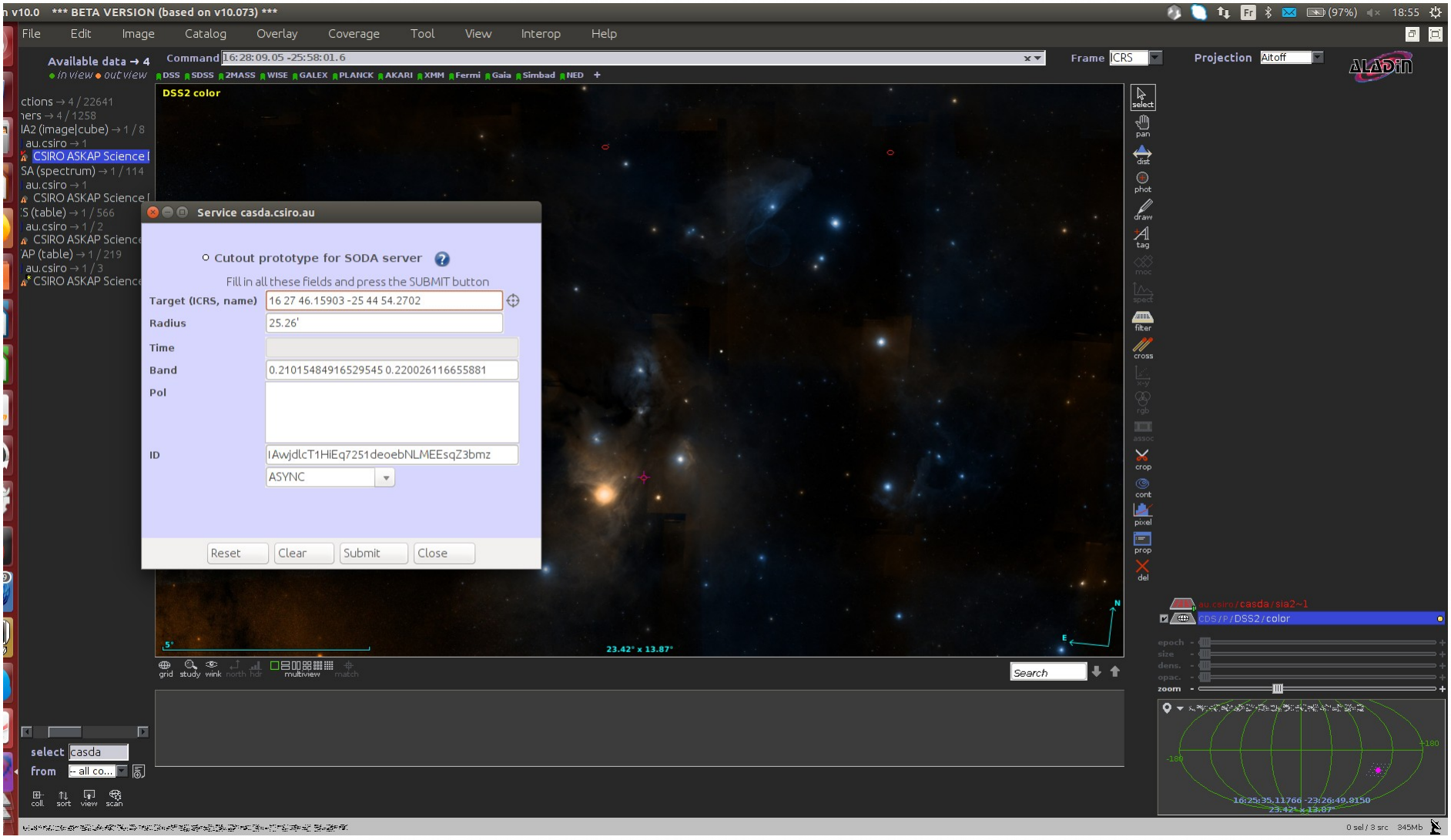
SODA asynchronous task completed



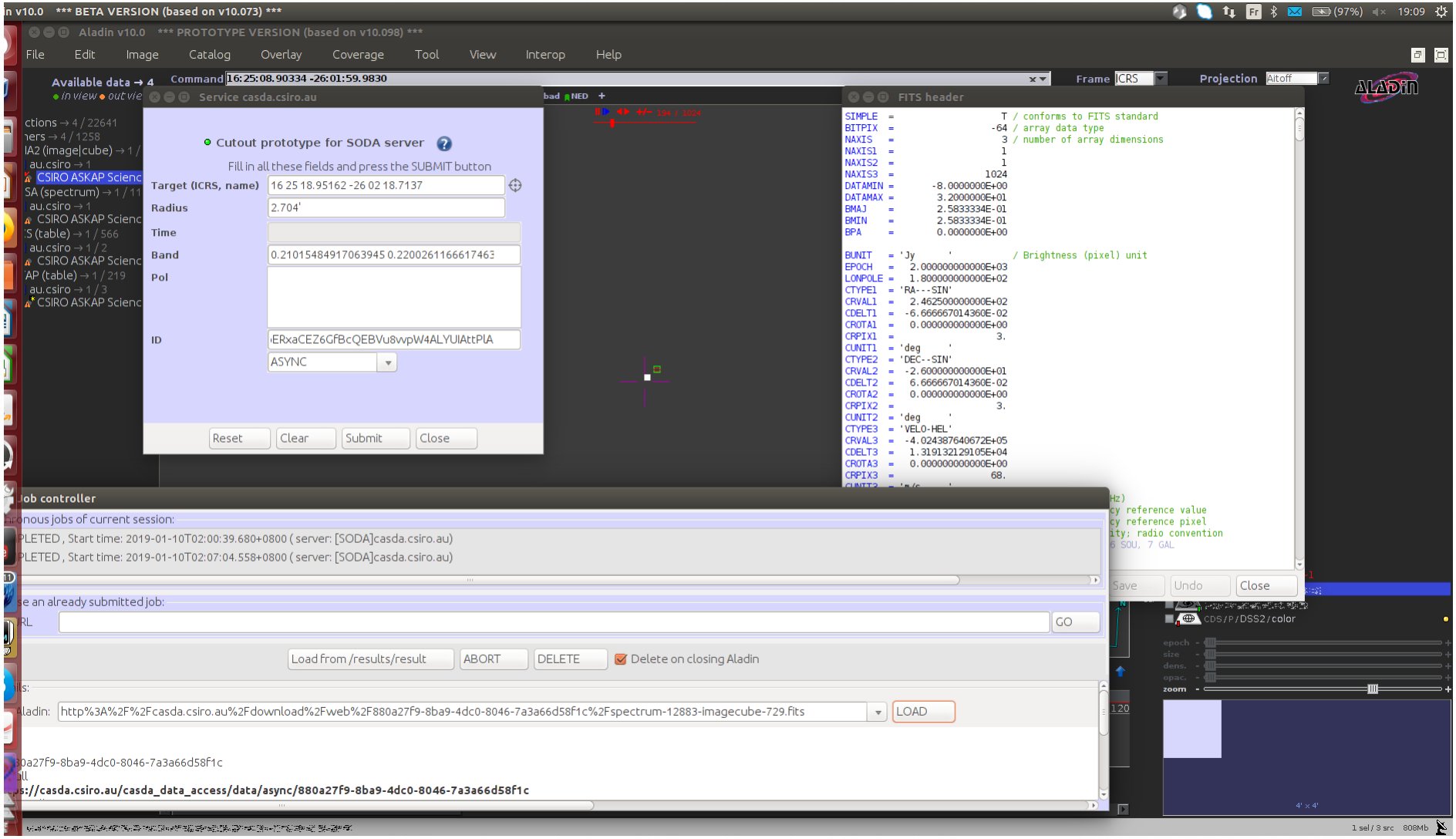
ASKAP data : CASDA services extracted subcube displayed



ASKAP data : CASDA services extracted spectrum



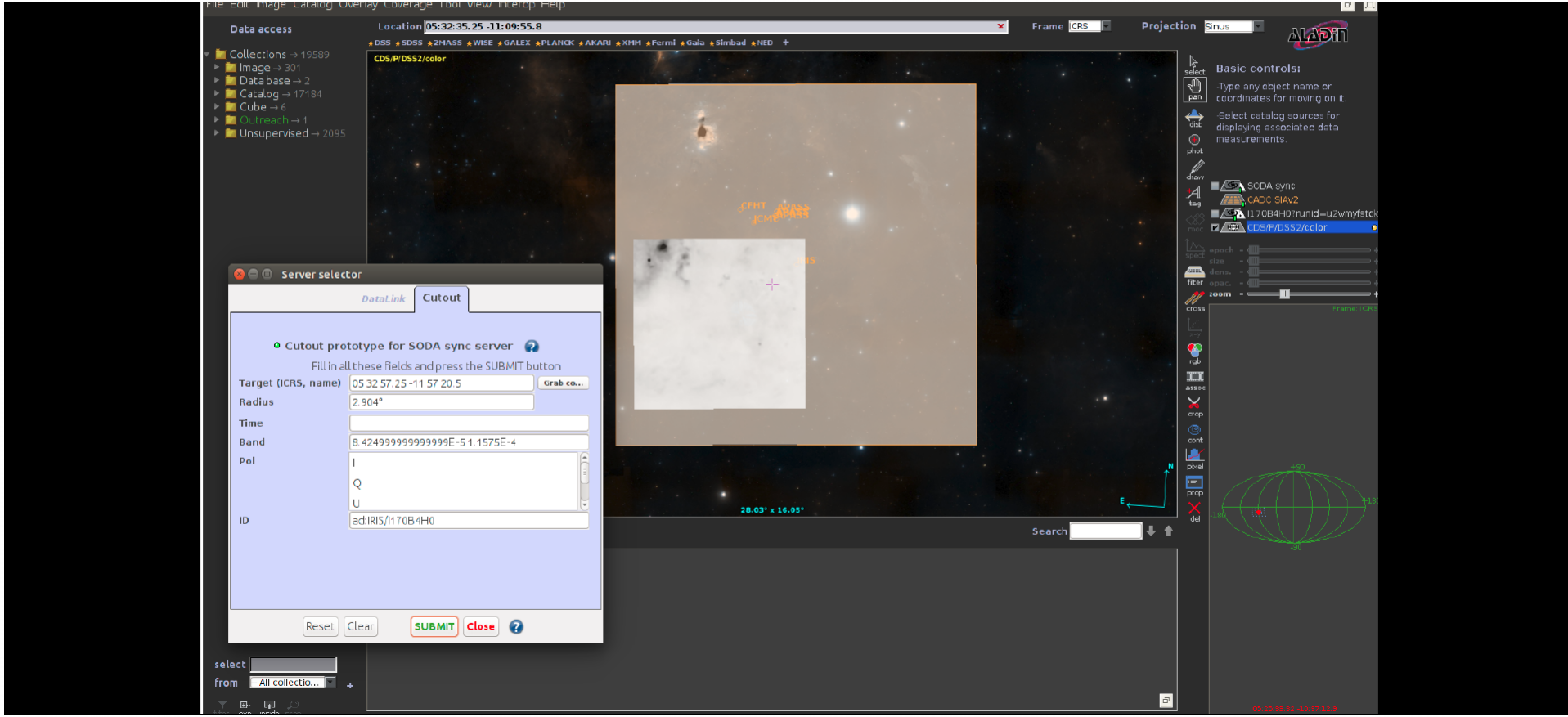
ASKAP data : CASDA services extracted spectrum



CADC and others

- CADC has a couple of Radio cube surveys such as
  - CGPS
  - ALMA
  - VLASS
- Plans at NRAO (collab with CADC?)
- Italian
- Others ? Sorry if any !!!

SIAV2 + DataLink + SODA at CADC



Time Domain ?

See Ada’s talk on Saturday

What else ?

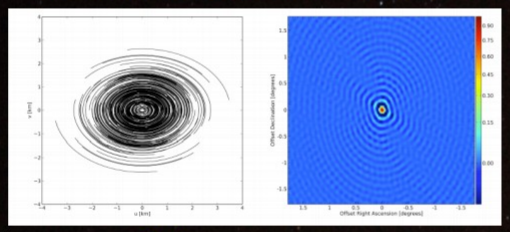
- A) extend « science ready data » availability in the VO .
  - Create new ObSTAP/SIAV2/DataLink/SODA services
  - Create HiPS for moment/continuum maps and access « original data »
- B) Visibility data (=VD) – what should be investigated/done :
  - Use ObsTAP/SIAP2 for VD rough discovery
  - VD fine grain characterisation and discovery
  - Integrate VD processing in VO landscape
  - VD VO representation/data modelling ?
- C) Others ? (Rotation measures, ...)

Visibility data?
(courtesy LOFAR/ASTRON partners)

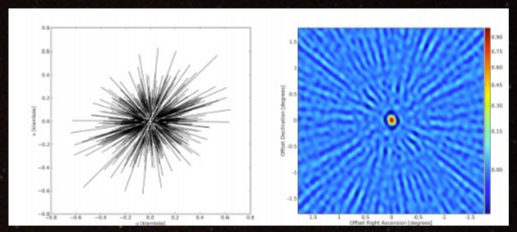
Data Types: Visibilities

ASTRON

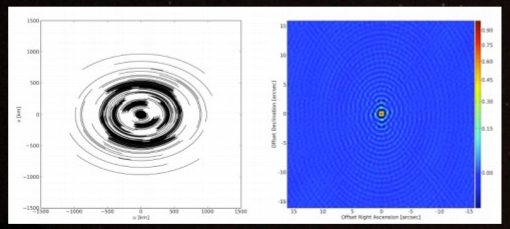
- Interferometer: Creating virtual aperture up to ~2000km (λ/D)
- Incomplete coverage
- Typical use of Radio Telescopes: imaging depending on science
- LOFAR: Projection effects (No moving dish/mirror)

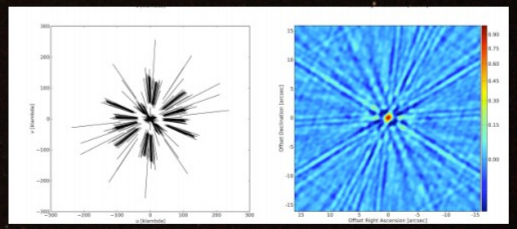


6 hours



1 second





Use ObsTAP/SIAP2 for VD rough discovery

- ObsCore has « visibility » dataproduct type
- Visibility data have rough spatial position/extension, time range, spectral range, polarisation

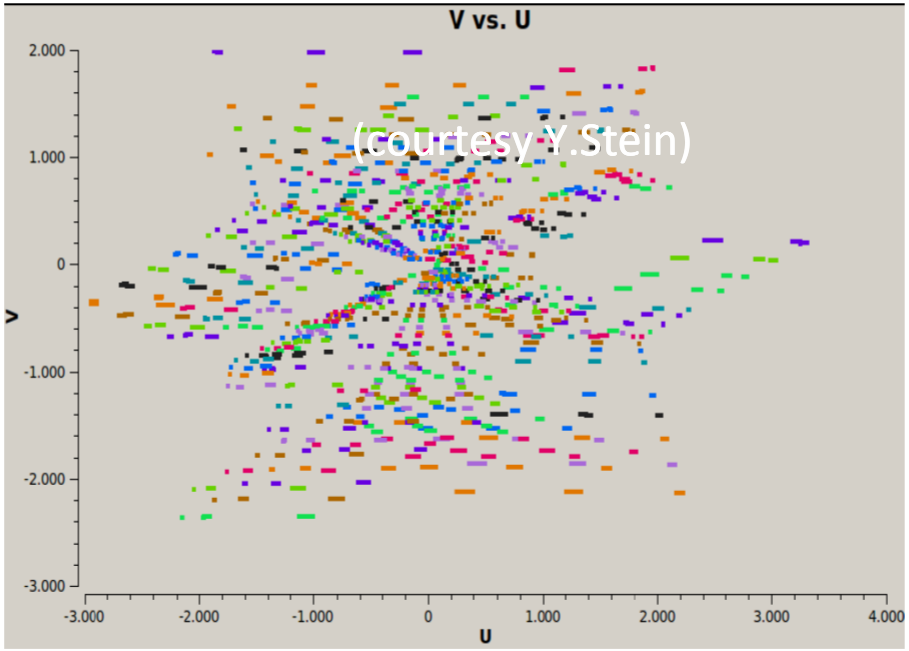
  → *roughly fit in ObsCore*
- Complex data structure can be searched via CAOM extensions ?

VD fine grain characterisation and discovery

- « coarse grain » discovery via ObsCore + DataLink for additional metadata
  - UV plots
  - frequency/amplitude or phase/amplitude plots
  - Beam size and shape
  - Number of antennas, instrumental details

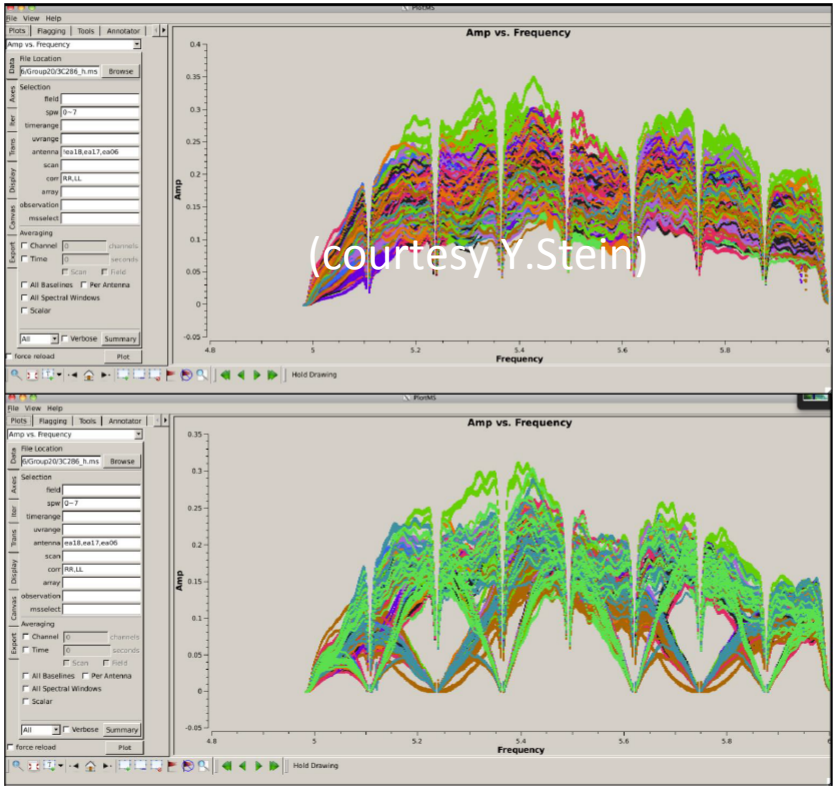
    → should we integrate this in « characterisation » DM ?
- Dynamic characterisation according to scientific goals : VO Web Service or VO application ?

Uv coverage



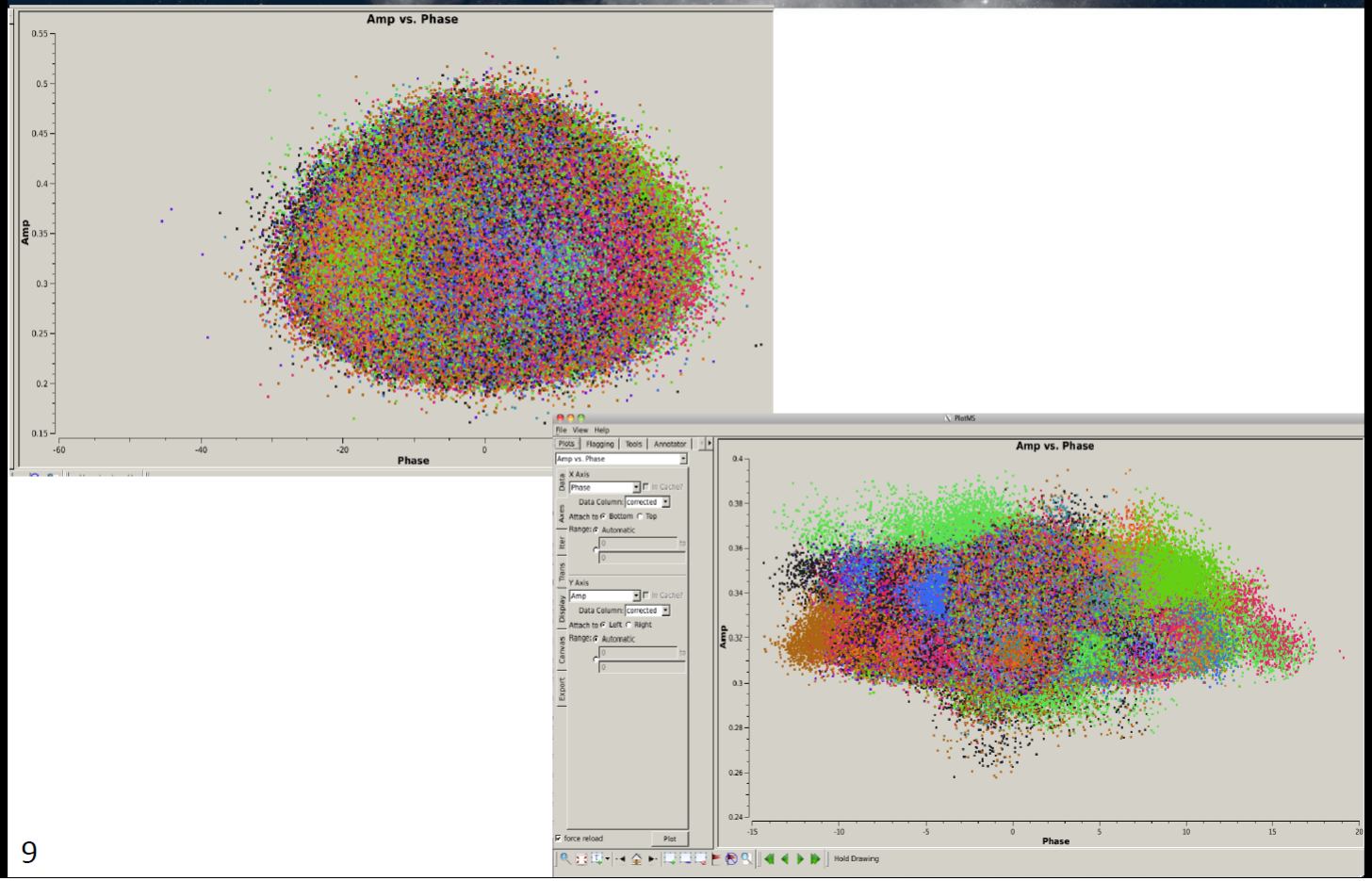
(courtesy Y.Stein)

Amp vs. Freq



(courtesy Y.Stein)

Amp vs. phase


Amp vs. Phase
Amp
Phase
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
-60
-40
-20
0
File View Help
Plots Flagging Tools Annotator
Amp vs. Phase
X Axis
Phase
Data Column: corrected
Attach to: Bottom Top
Range: Automatic
Y Axis
Amp
Data Column: corrected
Attach to: Left Right
Range: Automatic
force reload
Plot
Amp vs. Phase
Amp
Phase
0.4
0.38
0.36
0.34
0.32
0.3
0.28
0.26
0.24
-15
-10
-5
0
5
10
15
20
Hold Drawing

Integrate VD processing in VO landscape

- Use « provenance » to go back to original data
- Store and Make available science data reduced by individual scientists
- Propose a black box service providing « science ready data » according output characterisation
- Launch CASA code remotely close to the data
- Bring code to the data (docker/ jupiter notebook / platform)

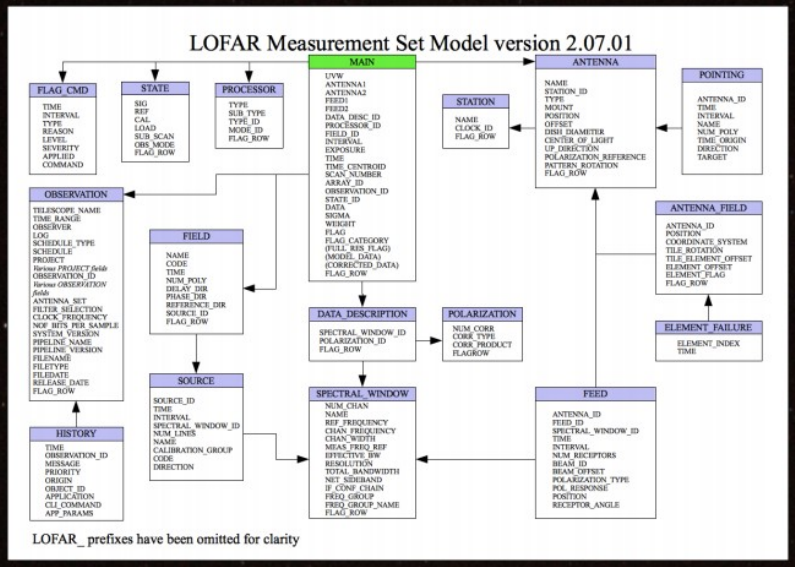
Visibility data modelling/representation

« measurement sets » are all what we need ?

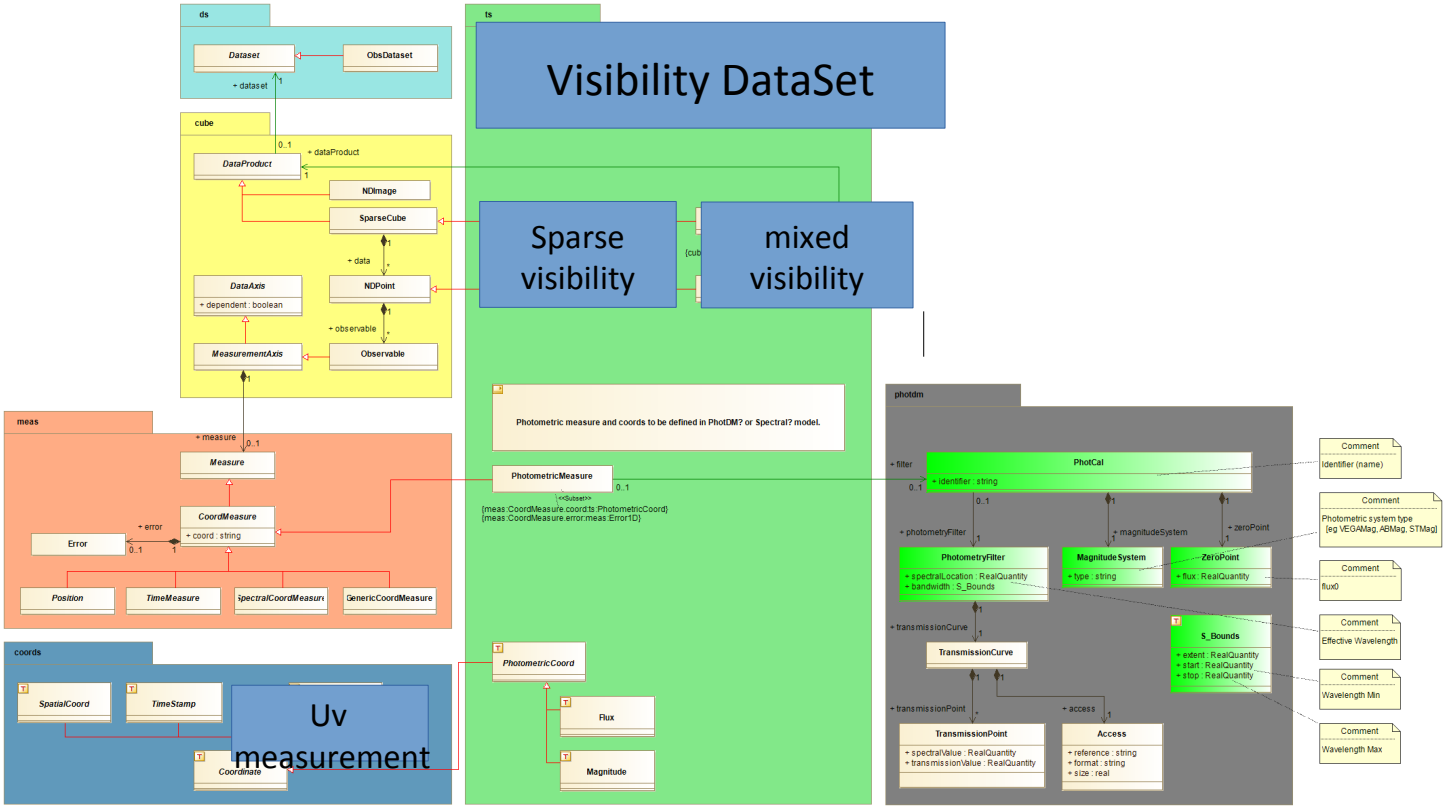
Data Types: Visibilities

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)



IVOA Visibility data model as an extension of CubeDM ?



Why : transform into time series, event lists, etc ? Combination with other spectral domains