The AENEAS survey of radio archives

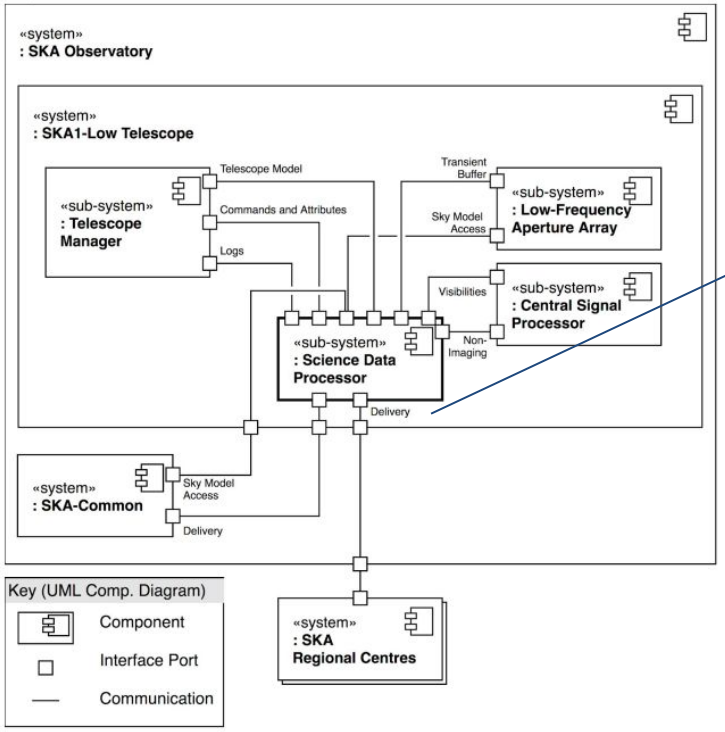
Use cases and user interface recommendations

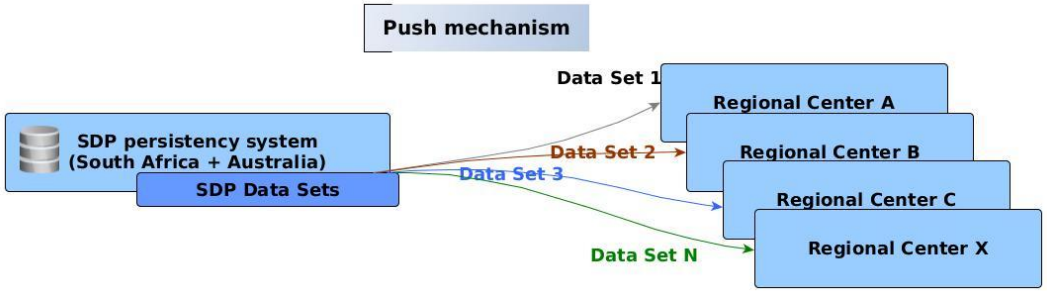
Outline

1. **The H2020 AENEAS project**
2. **Evolution of radio archives and use cases**
3. **Survey of radio archival interfaces**
4. **User interface recommendations**
5. **Conclusions**

AENEAS

AENEAS (**A**dvanced **E**uropean **N**etwork of **E**-infrastructures for **A**stronomy with the **S**KA) is an H2020 project established to design an **European SKA Data Centre (ESDC)** articulated in **SKA Regional Centres (ESRCs)**





AENEAS WP5

AENEAS (**A**dvanced **E**uropean **N**etwork of **E**-infrastructures for **A**stronomy with the **S**KA) is an H2020 project established to design an **European SKA Science Data Centre (ESDC)** articulated in **SKA Regional Centres (ESRCs)**

WP5 - Access and Knowledge Creation: design of the interface between ESDC and SKA

WP5 tasks:

WP5.1: Survey of existing user interaction models for large-scale radio astronomy facilities and integration of WP5 outputs into consolidated ESDC design study

WP5.2: Recommendations for the design of user interfaces for data discovery, access, and retrieval

WP5.3: Recommendations for the design of user interfaces for data processing, reprocessing, analysis, and visualization

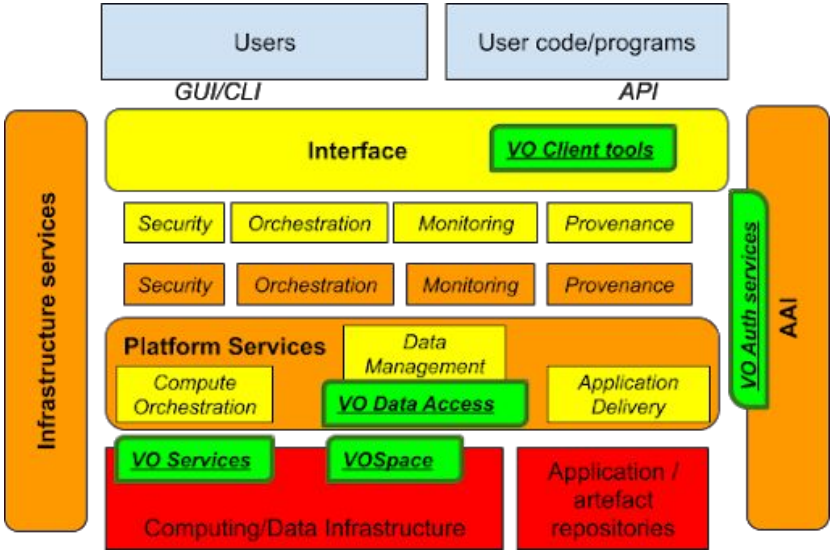
WP5.4: Integration with VO Interoperability Framework

WP5.5: Recommendations for the resourcing of an ESDC user interaction model

WP5.6: Recommendations for a plan of user community formation and knowledge distribution

AENEAS WP5 activities in a nutshell

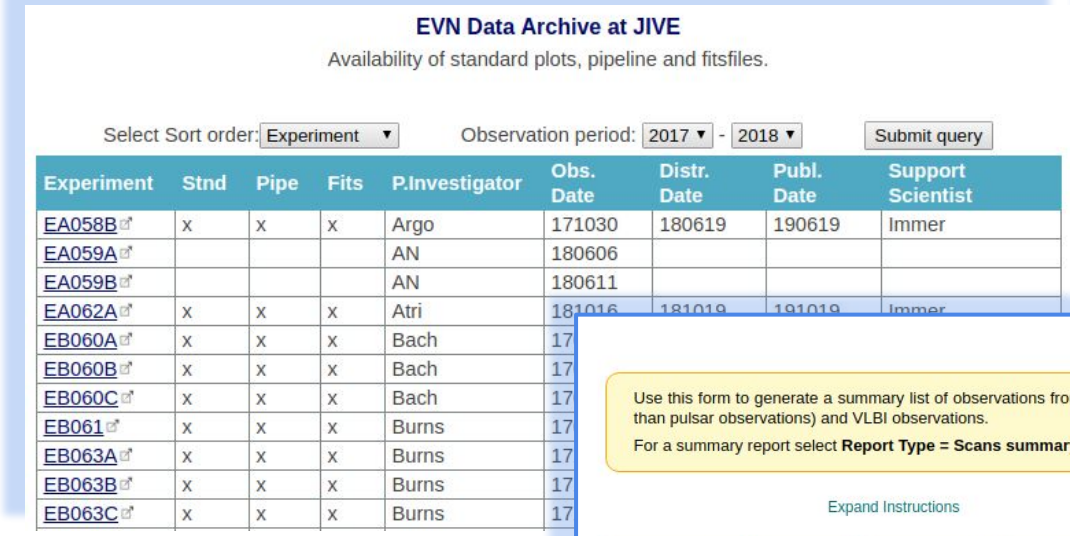
- **Survey of existing services and user needs**
- **Identification of gaps**
- **Deliver recommendations to fill such gaps** (www.aeneas2020.eu/project-deliverables), e.g.:

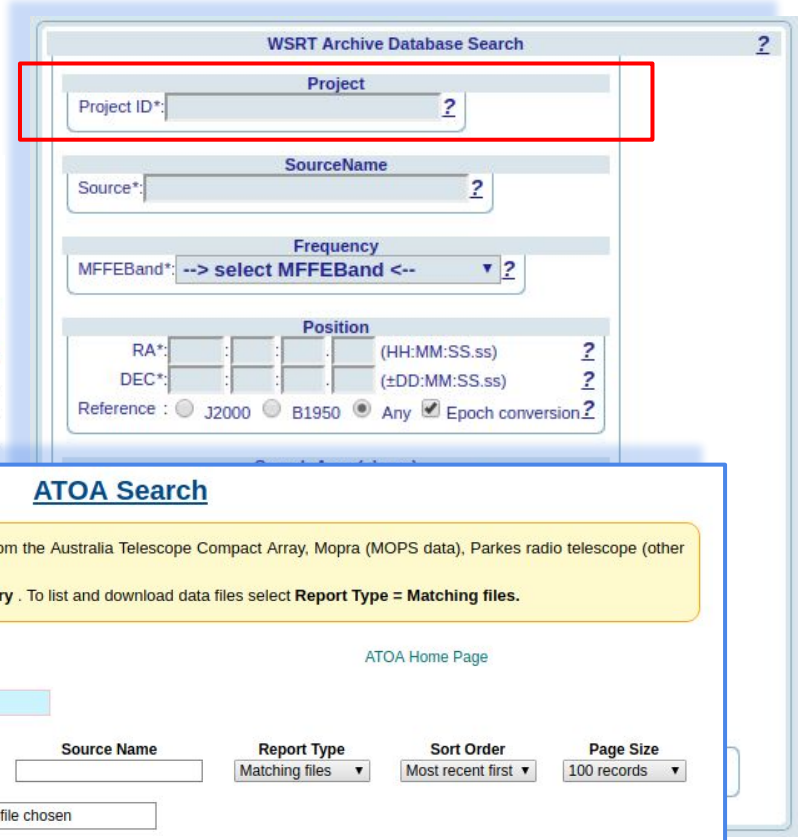


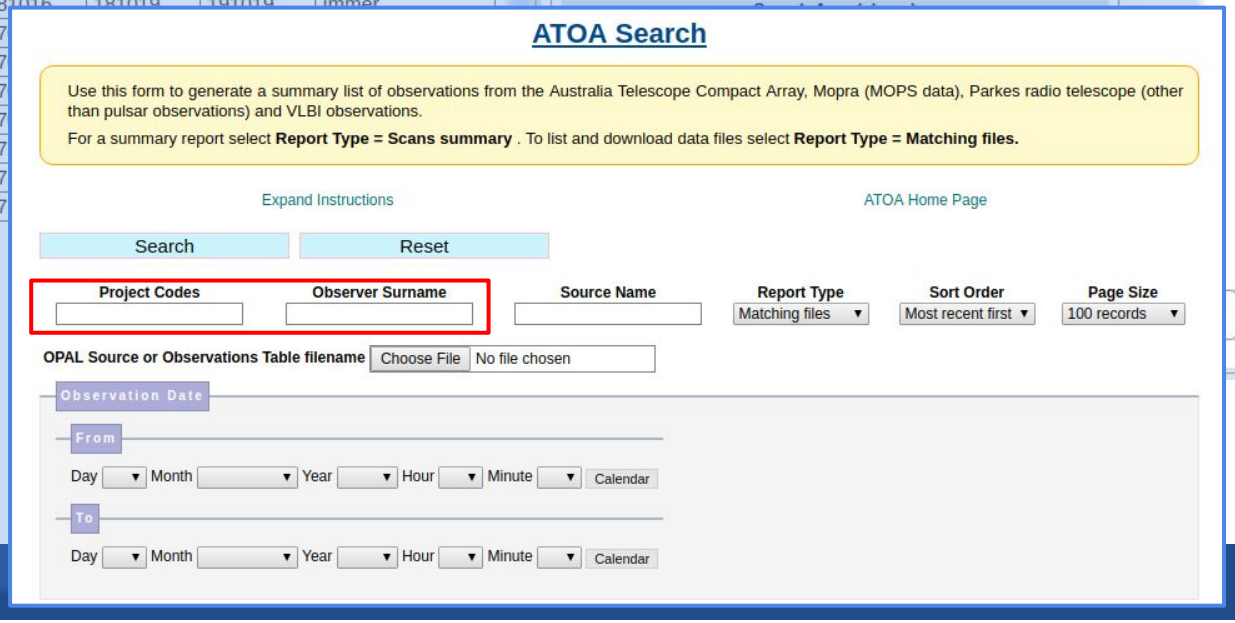
Science Gateway modular view, integrating VO services and VO-compatible tools (AENEAS-D5.5).

Evolution of the concept of archive

- **The archive is traditionally perceived as a mere repository of PI observations**







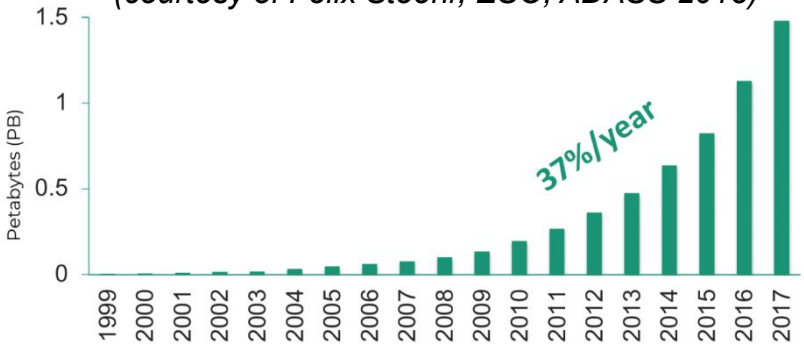
Evolution of the concept of archive

- The archive is traditionally perceived as a mere repository of PI observations
- **New telescopes (not only radio ones) already deliver huge datasets with a huge potential of reusability**

Energy, position, time, polarization, photon counts with unprecedented sensitivities

A single dataset might contain several objects not covered by the primary science goal of the observation (especially true for large FoV telescopes).

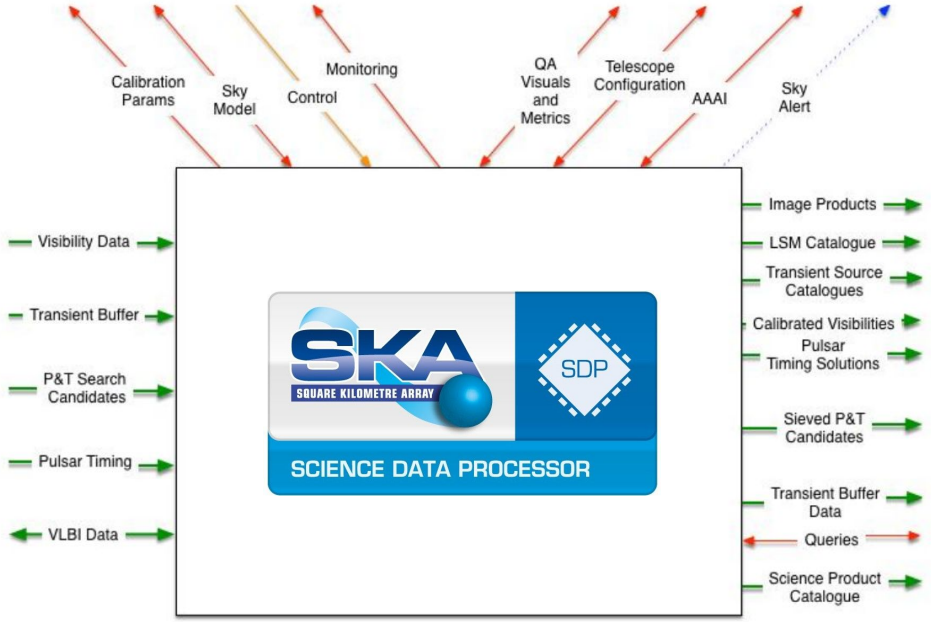
(courtesy of Felix Stoehr, ESO, ADASS 2018)

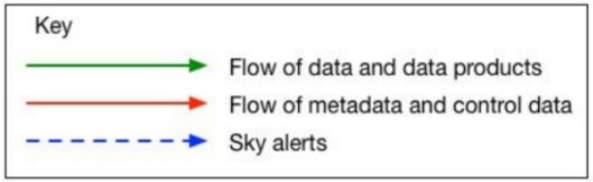


Telescope	Frequency (GHz)	FoV (sqdeg)	Raw data rate (PB/yr)
MWA	0.07-0.3	2500-200	3-8
LOFAR-LBA LOFAR-HBA	0.01-0.09 0.11-0.25	1700-7	7
ASKAP	0.7-1.8	30	70
SKA-LOW	0.05-0.35	30-4	4.9×10^{6}
SKA-MID	0.35-15.3	3.3-0.012	6.2×10^{4}

Evolution of the concept of archive *(in the SKA era)*

- **Raw data cannot be preserved** for a long time and data calibration/processing strongly couples with the archival system (e.g. for SKA SDP)





Data and metadata fluxes through SDP
(*cf.* Fig.3 of SKA1 SDP High Level Overview)

Evolution of the concept of archive *(in the SKA era)*

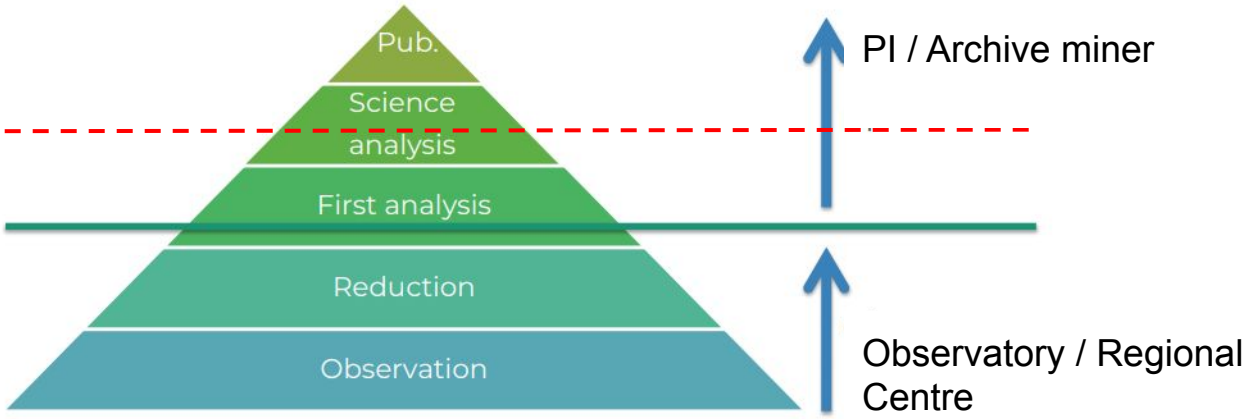
- **Raw data cannot be preserved** for a long time and data calibration/processing strongly couples with the archival system (e.g. for SKA SDP)
- Great boost in terms of the number people interested in data (**PIs from other research fields** and **archive miners**), hence a much more variegate community, asking for **different levels** of access and **different advanced data products**.

Evolution of the concept of archive *(in the SKA era)*

- **Raw data cannot be preserved** for a long time and data calibration/processing strongly couples with the archival system (e.g. for SKA SDP)
- Great boost in terms of the number people interested in data (**PIs from other research fields** and **archive miners**), hence a much more variegate community, asking for **different levels** of access and **different advanced data products**. E.g.:
  - **correlations** between radio properties and other **multi-wavelength/multi-messenge**r ones;
  - **population studies** by adopting **stacking techniques** (~millions of objects)
  - **cross-correlation** of radio objects with **completely different surveys** (e.g. the CMB, in order to constraint our cosmological model by looking at the Integrated Sachs-Wolfe effect)

Evolution of the concept of archive *(in the SKA era)*

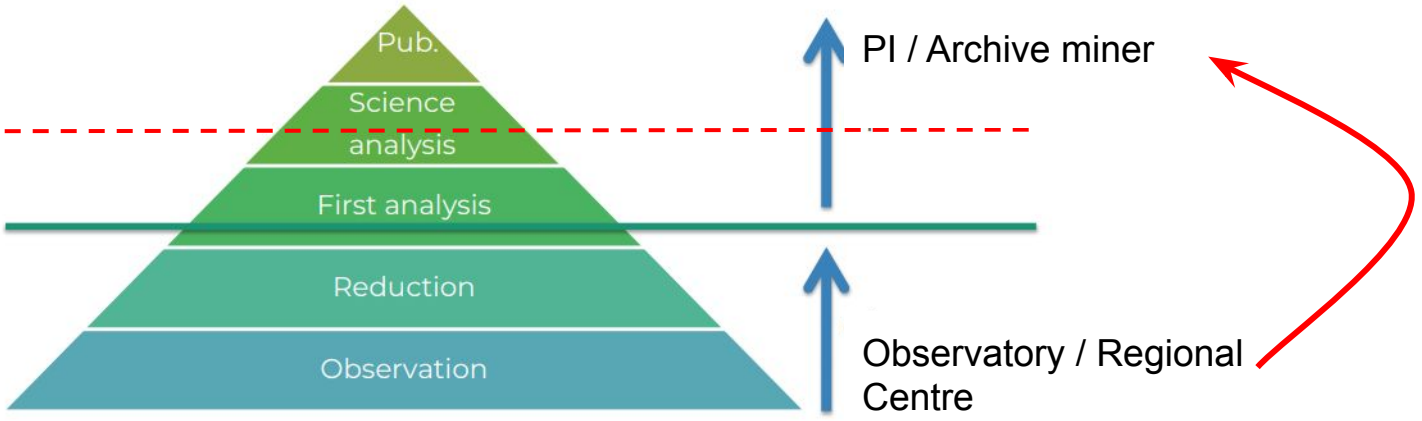
- **Raw data cannot be preserved** for a long time and data calibration/processing strongly couples with the archival system (e.g. for SKA SDP)
- Great boost in terms of the number people interested in data (**PIs from other research fields** and **archive miners**), hence a much more variegate community, asking for **different levels** of access and **different advanced data products**.
- **More responsibility** for the observatory and/or the infrastructure (e.g. Regional Centres)



Courtesy of Felix Stoehr (ESO, ADASS 2018)

Evolution of the concept of archive *(in the SKA era)*

- **Raw data cannot be preserved** for a long time and data calibration/processing strongly couples with the archival system (e.g. for SKA SDP)
- Great boost in terms of the number people interested in data (**PIs from other research fields** and **archive miners**), hence a much more variegate community, asking for **different levels** of access and **different advanced data products**.
- **More responsibility** for the observatory and/or the infrastructure (e.g. Regional Centres)



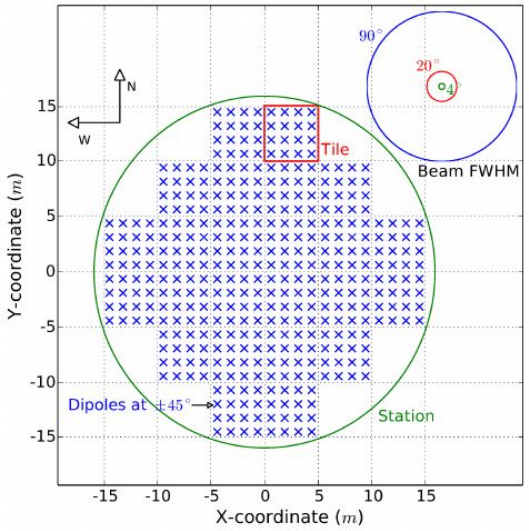
Courtesy of Felix Stoehr (ESO, ADASS 2018)

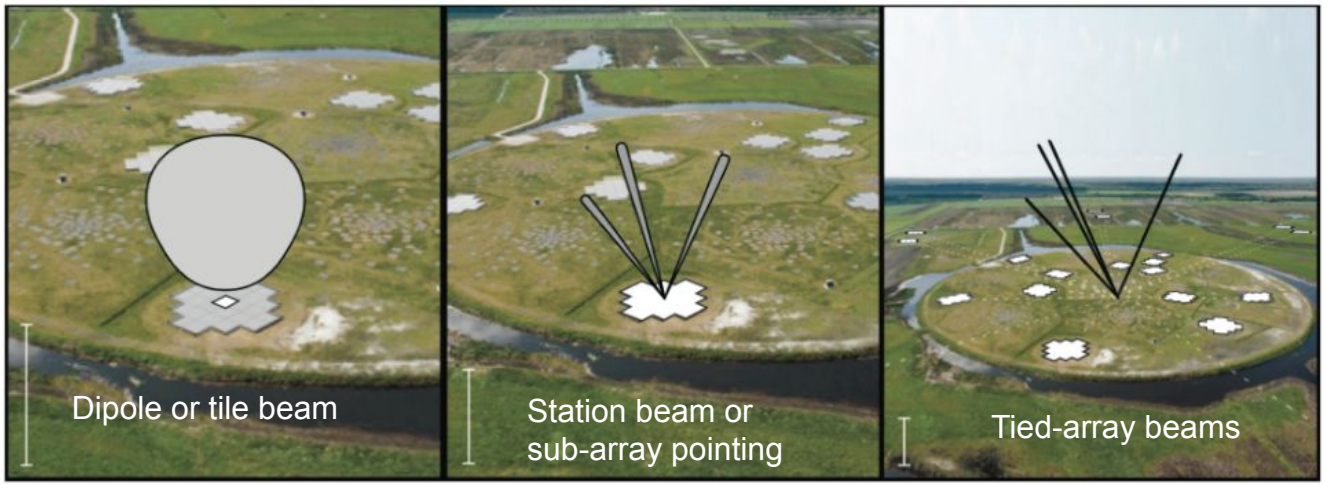
A survey of radio archives *(... and not only)*

- We focused on the most important SKA *precursors*: ASKAP, ***MWA***, (KAT-7 and MeerKAT)
- … and *pathfinders*: ***LOFAR***, EVN, GMRT, e-MERLIN, ***(J)VLA***
- We also considered: WSRT, ATCA, ***VLBA (and GBT)***, ***ALMA***
- We extended the analysis to: CADC, ESO Science Archive, CDS, IA2

LOFAR

- The **Lo**w-**F**requency **AR**ray is operated by ASTRON and observes between 10 and 250 MHz (LBA: 10-90 MHz; HBA: 110-250 MHz)
- It consists of 24 core stations, 14 remote stations and 13 international ones *(... so far)*.
- It can collect data through three modes of acquisition:
  - interferometry, i.e. correlated visibilities
  - beam-formation *(Coherent Stokes, Incoherent Stokes, Fly's Eye)*
  - transient buffer modes
- Time/spectral resolution can be traded in order to fit with scientific goals





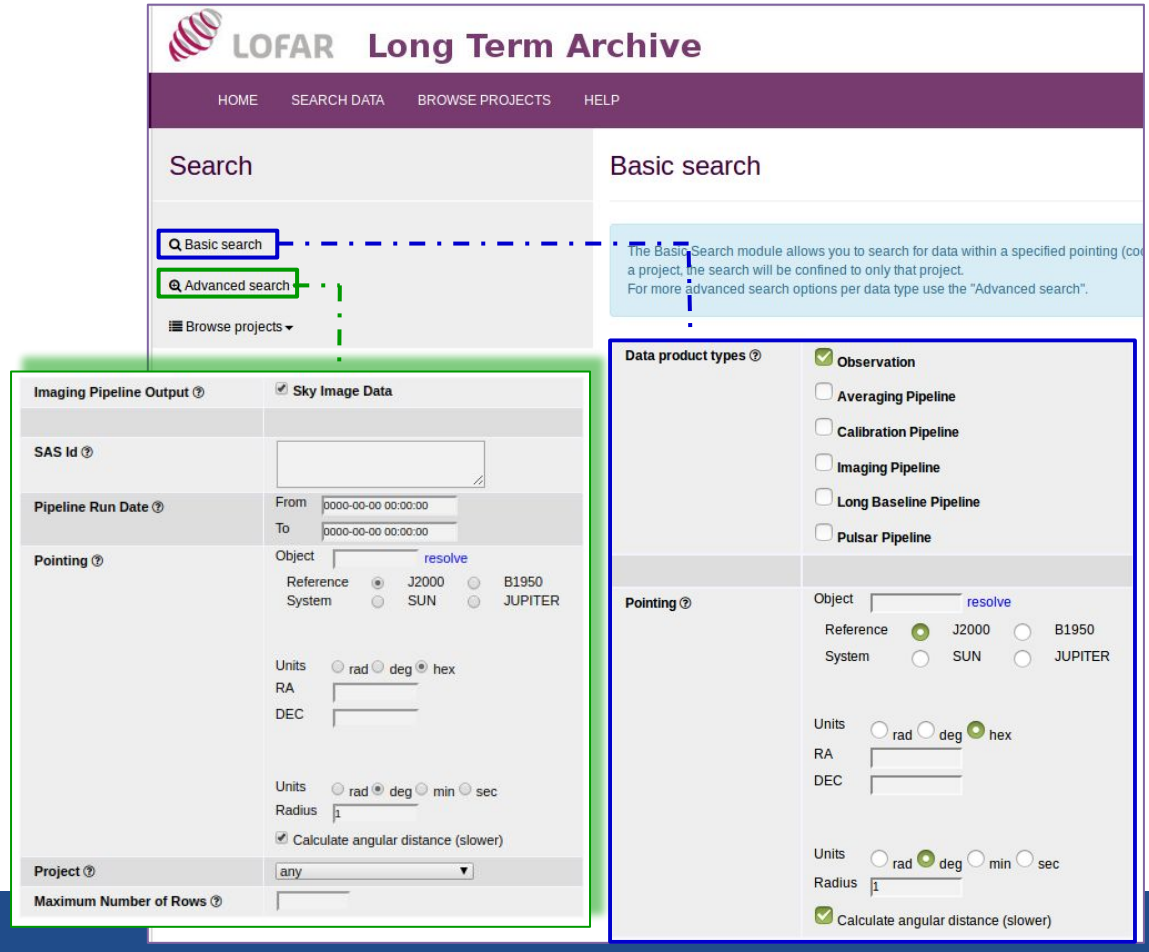
LOFAR LTA

- The LOFAR **L**ong **T**erm **A**rchive (LTA) allows metadata search without any account (you need one for downloading/staging requests)

- It can be decomposed in two sections: left panel, namely **basic search**, **advanced search** and **browse project**; the right panel displaying a dedicated search mask.

- There is also a server for **SQL queries** and a **Python interface**, as well as a **Python API** for improving user control over staging requests.

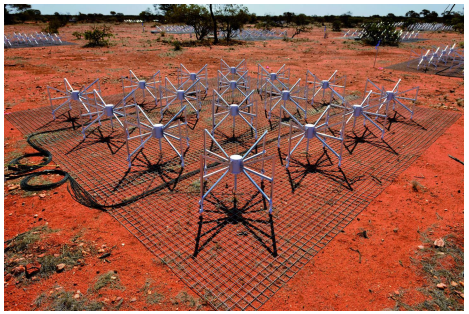
GUI of LOFAR LTA (https://lta.lofar.eu/) *(basic and advanced search are shown)*



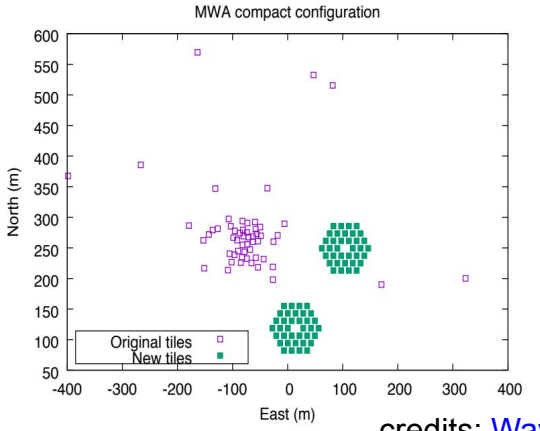
MWA

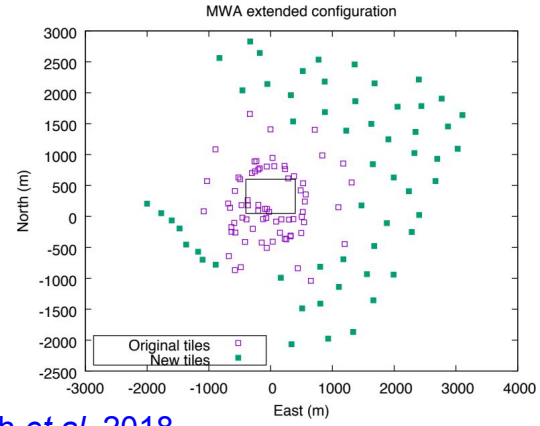
- The **M**urchison **W**idefield **A**rray is run by a consortium of **21 institutions** from Australia, New Zealand, China, Japan, Canada and USA, led by the **Curtin University**.
- It consists of 256 stations (tiles), 128 (chosen among two configurations, can be correlated at a time)

- Main features:
  - wide field of view (200-2500 sqdeg)
  - instantaneous bandwidth of 30.72MHz
  - spectral resolution of 20 kHz
  - temporal resolution of 0.5s
  - wide frequency range (70–300 MHz) with flexible tuning
  - digital design with extreme frequency and pointing agility, wide fractional bandwidths and considerable signal processing capabilities.



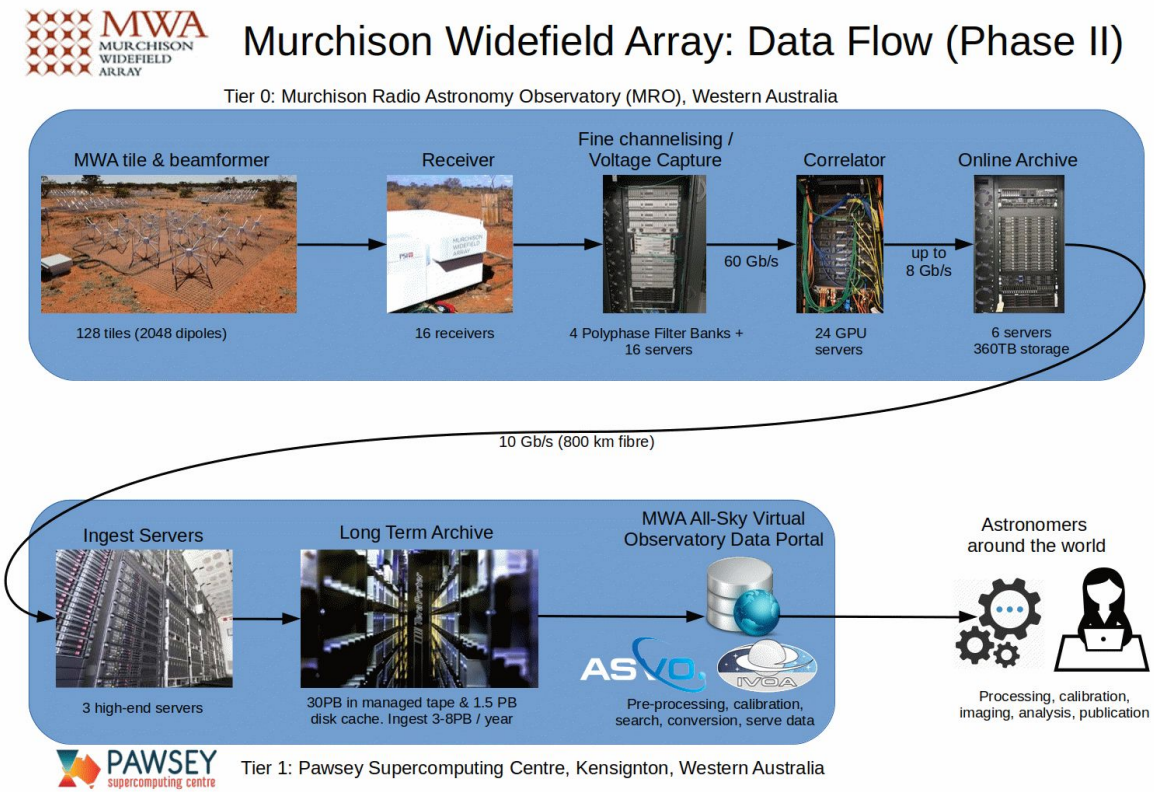
An MWA tile (4 x 4 dipoles)





credits: Wayth *et al.* 2018

MWA Archive



High level view of the MWA dataflow (courtesy of Greg Sleap, ICRAR)

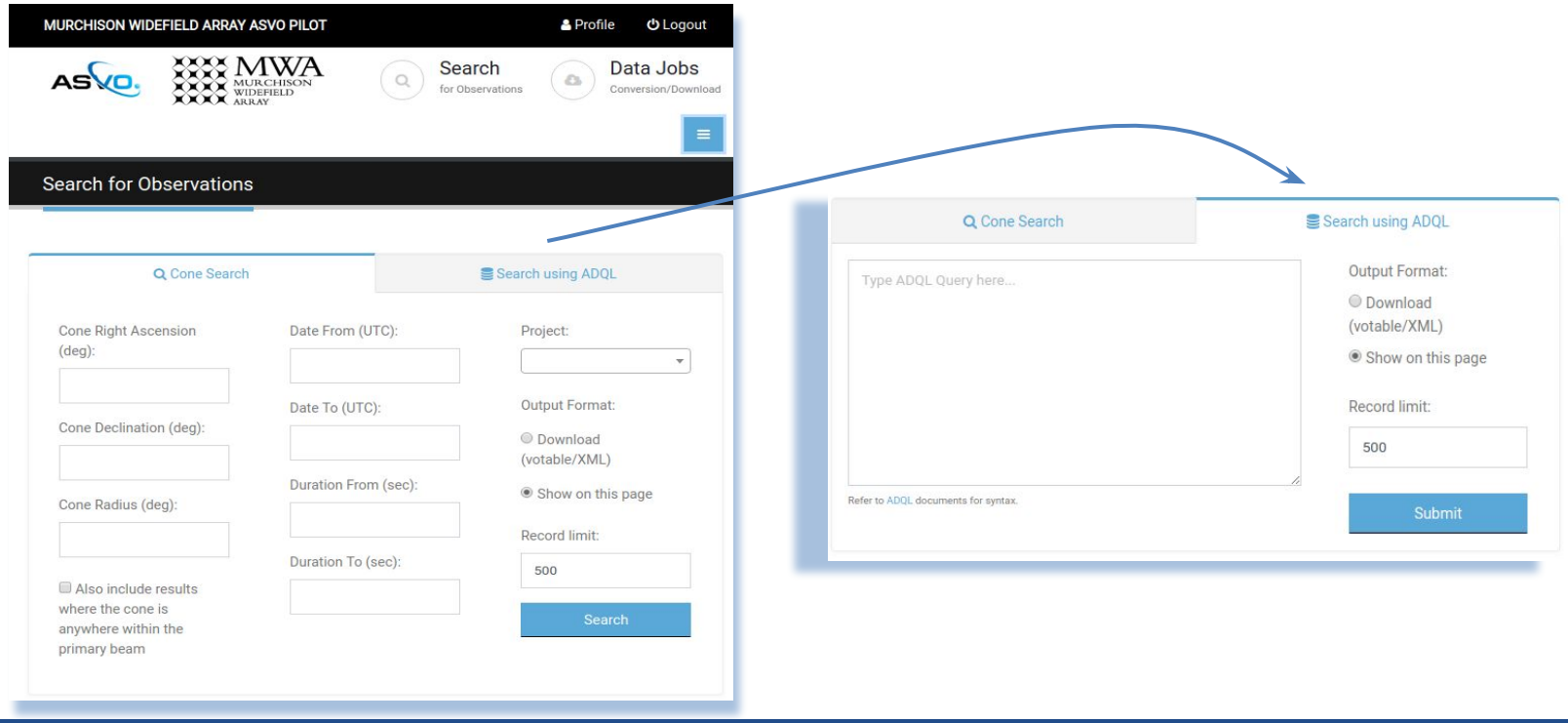
The **archive core** is constituted by the *Next-Generation Archive System* (NGAS)

Tier 2 mirrored archives at MIT, Victoria University of Wellington, Raman Research Institute, University of Melbourne, and in Trieste.

Data are carried over by the Australia *Academic and Research Network* (AARNET) across the Pacific Ocean

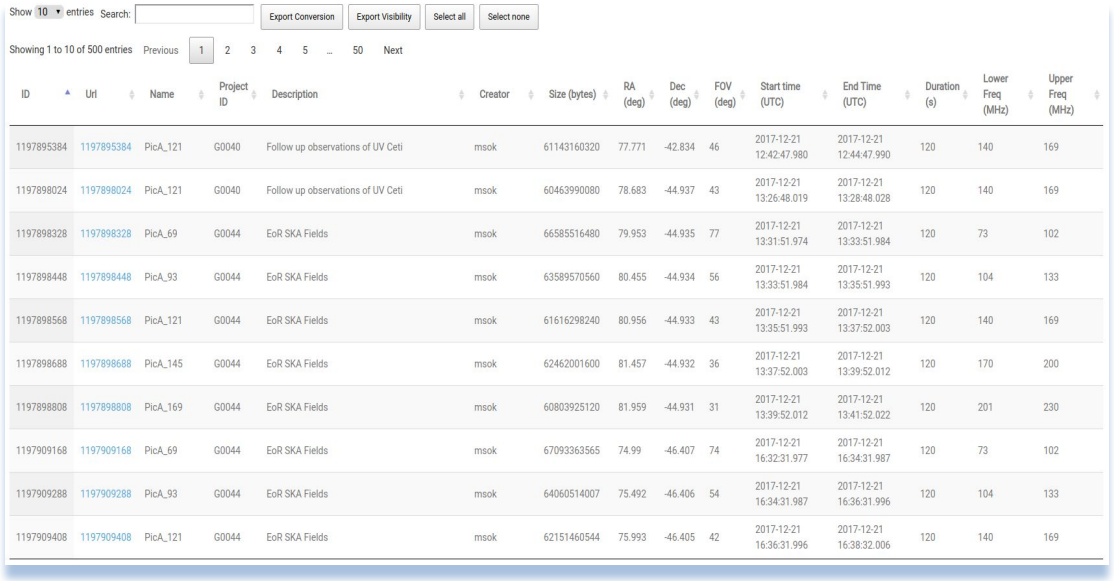
MWA Archive

- Access to the MWA Archive is granted by the *MWA All-Sky Virtual Observatory* (ASVO) data portal, through: a web dashboard, a Python API (Manta-ray-client) or a VO TAP service.
- It is requested a **registration on the ASVO portal** to use the web dashboard or there is the possibility to exploit **eduGain** (ORCID and United ID are also supported).



MWA Archive

- The Data Export Tool allows to perform some **pre-calibration tasks** and **data format conversions**

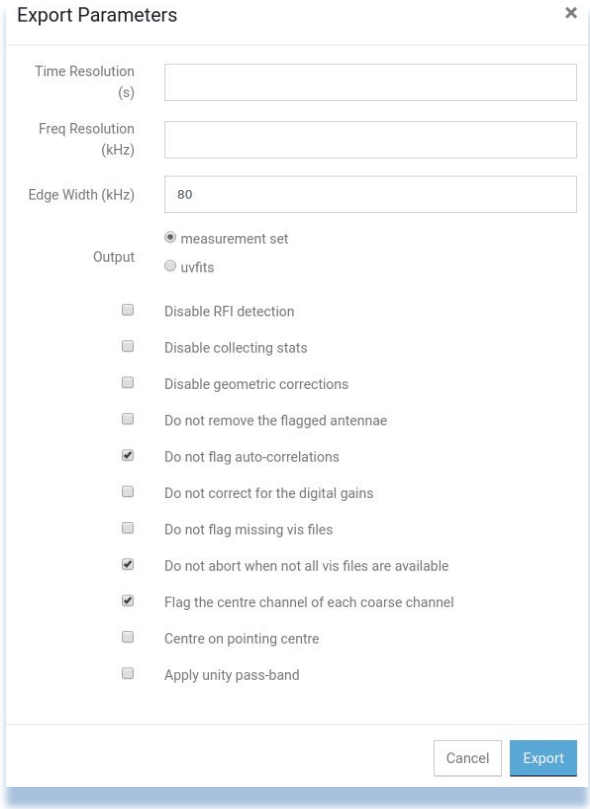


Show 10 entries Search: Export Conversion Export Visibility Select all Select none

Showing 1 to 10 of 500 entries Previous 1 2 3 4 5 … 50 Next

ID	Url	Name	Project ID	Description	Creator	Size (bytes)	RA (deg)	Dec (deg)	FOV (deg)	Start time (UTC)	End Time (UTC)	Duration (s)	Lower Freq (MHz)	Upper Freq (MHz)
1197895384	1197895384	PicA_121	G0040	Follow up observations of UV Ceti	msok	61143160320	77.771	-42.834	46	2017-12-21 12:42:47.980	2017-12-21 12:44:47.990	120	140	169
1197898024	1197898024	PicA_121	G0040	Follow up observations of UV Ceti	msok	60463990080	78.683	-44.937	43	2017-12-21 13:26:48.019	2017-12-21 13:28:48.028	120	140	169
1197898328	1197898328	PicA_69	G0044	EoR SKA Fields	msok	66585516480	79.953	-44.935	77	2017-12-21 13:31:51.974	2017-12-21 13:33:51.984	120	73	102
1197898448	1197898448	PicA_93	G0044	EoR SKA Fields	msok	63589570560	80.455	-44.934	56	2017-12-21 13:33:51.984	2017-12-21 13:35:51.993	120	104	133
1197898568	1197898568	PicA_121	G0044	EoR SKA Fields	msok	61616298240	80.956	-44.933	43	2017-12-21 13:35:51.993	2017-12-21 13:37:52.003	120	140	169
1197898688	1197898688	PicA_145	G0044	EoR SKA Fields	msok	62462001600	81.457	-44.932	36	2017-12-21 13:37:52.003	2017-12-21 13:39:52.012	120	170	200
1197898808	1197898808	PicA_169	G0044	EoR SKA Fields	msok	60803925120	81.959	-44.931	31	2017-12-21 13:39:52.012	2017-12-21 13:41:52.022	120	201	230
1197909168	1197909168	PicA_69	G0044	EoR SKA Fields	msok	67093363565	74.99	-46.407	74	2017-12-21 16:32:31.977	2017-12-21 16:34:31.987	120	73	102
1197909288	1197909288	PicA_93	G0044	EoR SKA Fields	msok	64060514007	75.492	-46.406	54	2017-12-21 16:34:31.987	2017-12-21 16:36:31.996	120	104	133
1197909408	1197909408	PicA_121	G0044	EoR SKA Fields	msok	62151460544	75.993	-46.405	42	2017-12-21 16:36:31.996	2017-12-21 16:38:32.006	120	140	169

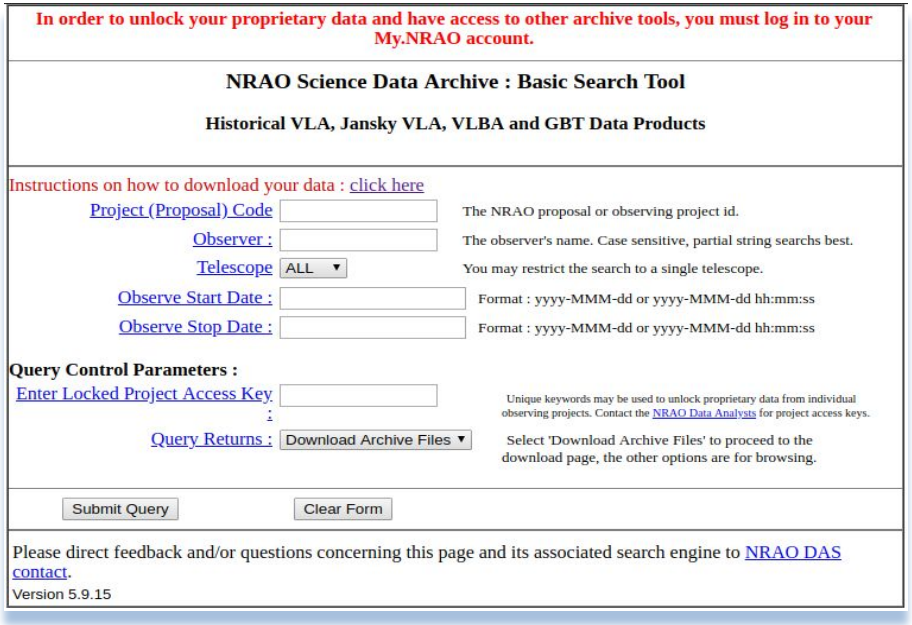
An example of web returned results for a cone search of 10 arcsec around the core of Pictor A

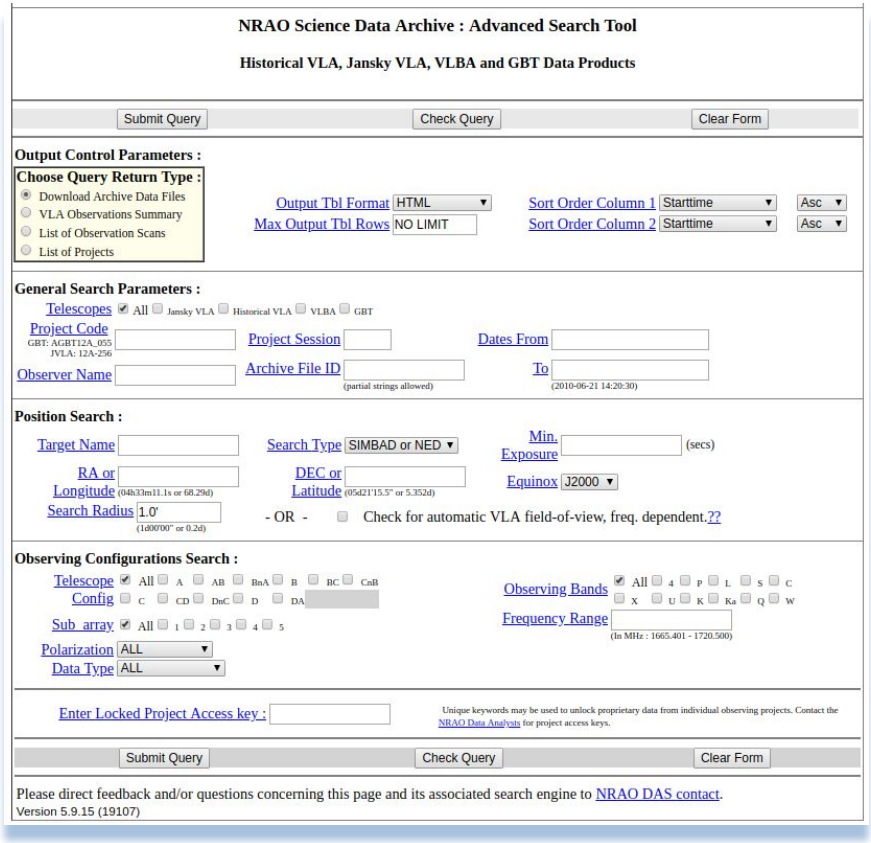


The Export Conversion web interface for selected MWA data

The new NRAO portal

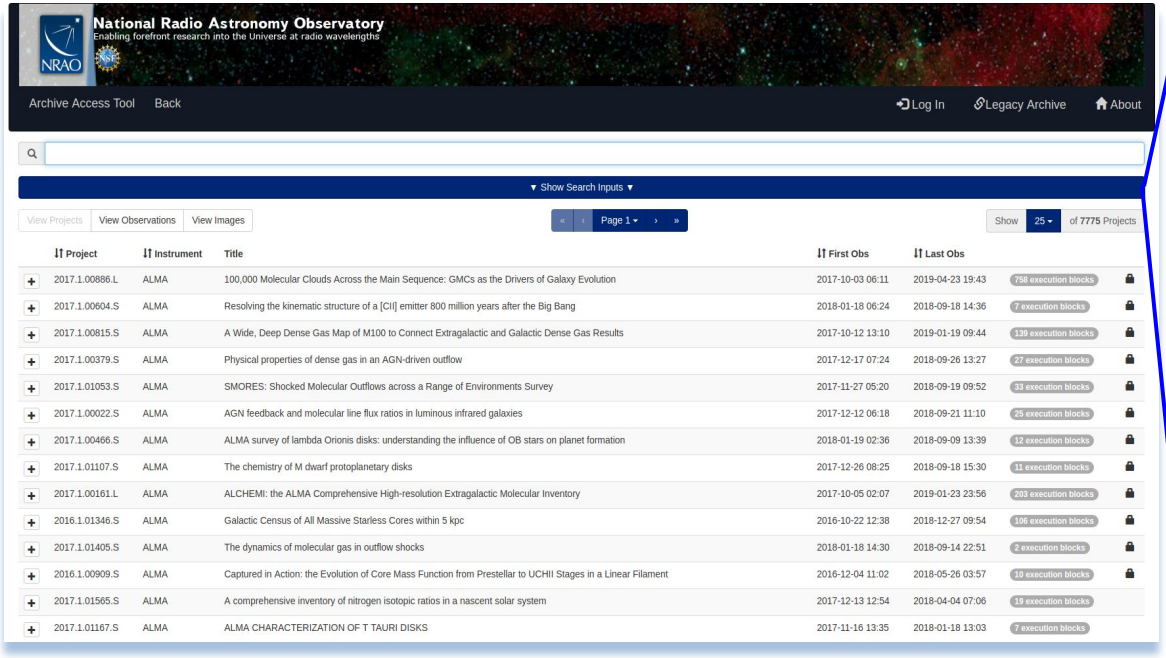
- The traditional *NRAO Data Archive System* (now called NRAO Legacy Archive) used to be the unified data archive for (J)VLA, VLBA and GBT data, and it is going to be completely replaced

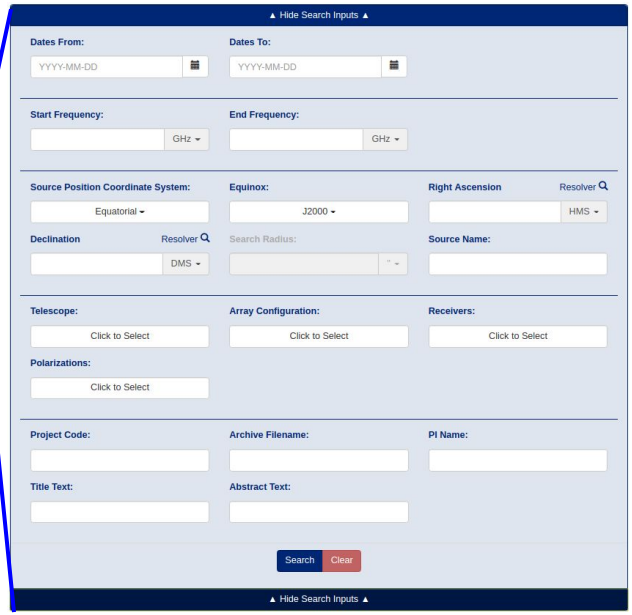




The new NRAO portal

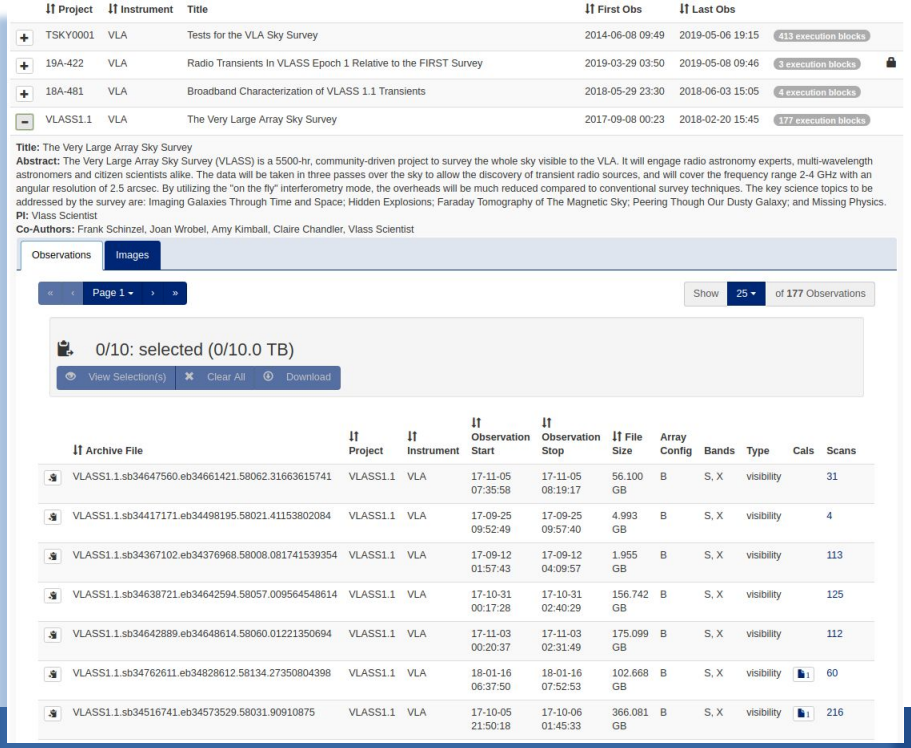
- The new NRAO Science Data Archive is still under development but has been already publicly released: also ALMA data are being ingested.
- Major features: **more responsive**, **interactive and significantly cleaner** with respect to its predecessor
- Access to raw observations and images via a new **Archive Access Tool (AAT)** and to processing via the Pipeline Processing Interface (PPI)

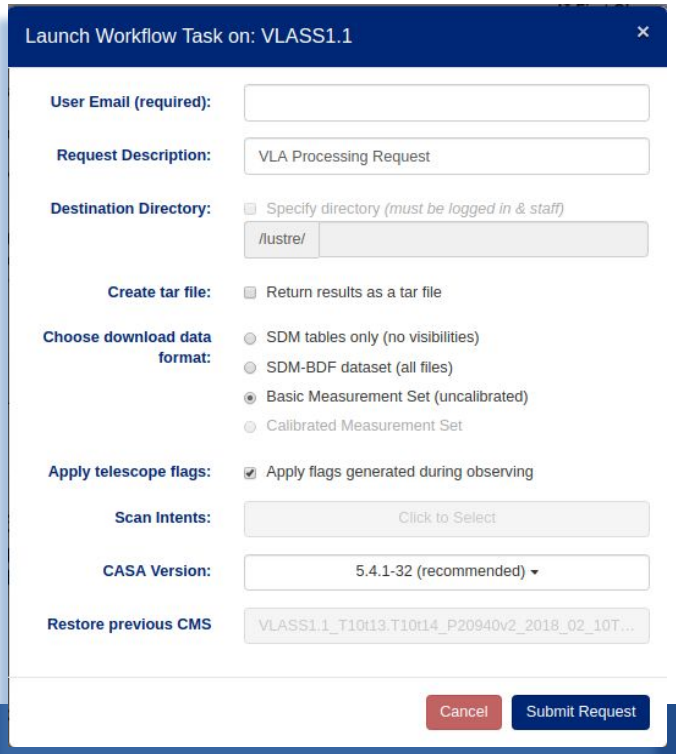




The new NRAO portal

- The new [NRAO Science Data Archive](#) is still under development but has been already publicly released: also ALMA data are being ingested.
- Major features: **more responsive, interactive and significantly cleaner** with respect to its predecessor
- Access to raw observations and images via a new Archive Access Tool (AAT) and to processing via the **Pipeline Processing Interface (PPI)**





Recommendations for the SKA Archive

The **SKA Science Archive** will contain not only **observatory products** but also **metadata**, **advanced data products** and **codes** needed to generate the latter.

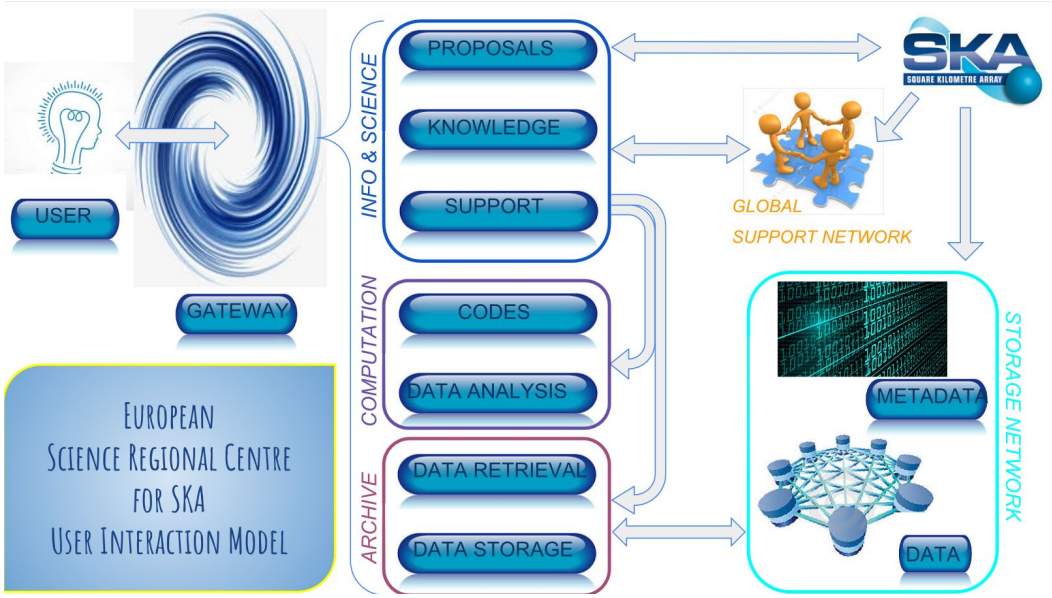
- We recommend a **multi-mask archival interface** (e.g. a basic search for observatory products and associated metadata, a dedicated or advanced mask for each product type
- **Observatory products, advanced products and codes** should be **clearly linked**
- **Proposal metadata** should be accessible
- The **query interface** should be **well documented**, with **different levels of access** (from a simple keyword-based GUI to programmatic/via tools)
- **Query refinement**, e.g. in terms of sensitivity and quality metrics, as well as lower resolution previews
- **VLBI archives** should be **homogenized** for SKA products
- **Data cutout services** and, more generally, exploitation of **VO services**
- The archive should be accessible through a **common gateway** to any user in the world, **managed at SRC level**, but on a **shared effort over the international partnership**

Conclusions

The SKA is a **strategic research infrastructure** which poses challenges at different levels:

- **the boundary** between the SKAO and the SKA Data Center
- **huge amounts of data** which require unprecedented computational power (distributed resources needed) processed up to science-ready products, and the need **of interoperating** these with different wavelengths/messengers datasets
- define **different roles** and find **human resources** for such efforts

The AENEAS project has already provided a **user interaction model** for the European SKA Data Centre (ESDC) articulated in **European SKA Regional Centre (ESRC) nodes**



courtesy of Marcella Massardi - AENEAS WP5 leader -