Radio data: the INAF experience and new challenges

Alessandra Zanichelli

INAF – Istituto di Radioastronomia

Outline

- INAF and the radio astronomical facilities: international and national landscape
- The Italian radio telescopes: a coordinated network
- The INAF radio data Archive
- Challenges for a modern archive in the epoch of big data

The international landscape

INAF through UTG2 involved in many **international** radio infrastructures.
(See talks by Smareglia, Botteon, Taffoni, Massardi)

The international landscape

INAF through UTG2 involved in many **international** radio infrastructures.
(See talks by Smareglia, Botteon, Taffoni, Massardi)

SKA and pathfinders/precursors

The international landscape

INAF through UTG2 involved in many **international** radio infrastructures (See talks by Smareglia, Botteon, Taffoni, Massardi)

SKA and pathfinders/precursors

Contribution to data archiving: MWA @IA2

The international landscape

INAF through UTG2 involved in many **international** radio infrastructures (See talks by Smareglia, Botteon, Taffoni, Massardi)

SKA and pathfinders/precursors

LOFAR and LOFAR.IT

Short-term archive already operational

LOFAR.IT: standalone mode 10% of the time.

The international landscape

INAF through UTG2 involved in many **international** radio infrastructures (See talks by Smareglia, Botteon, Taffoni, Massardi)

ALMA and ARI-L

archival image re-processing

The national landscape

Space weather: the first example of an INAF radio data archive?

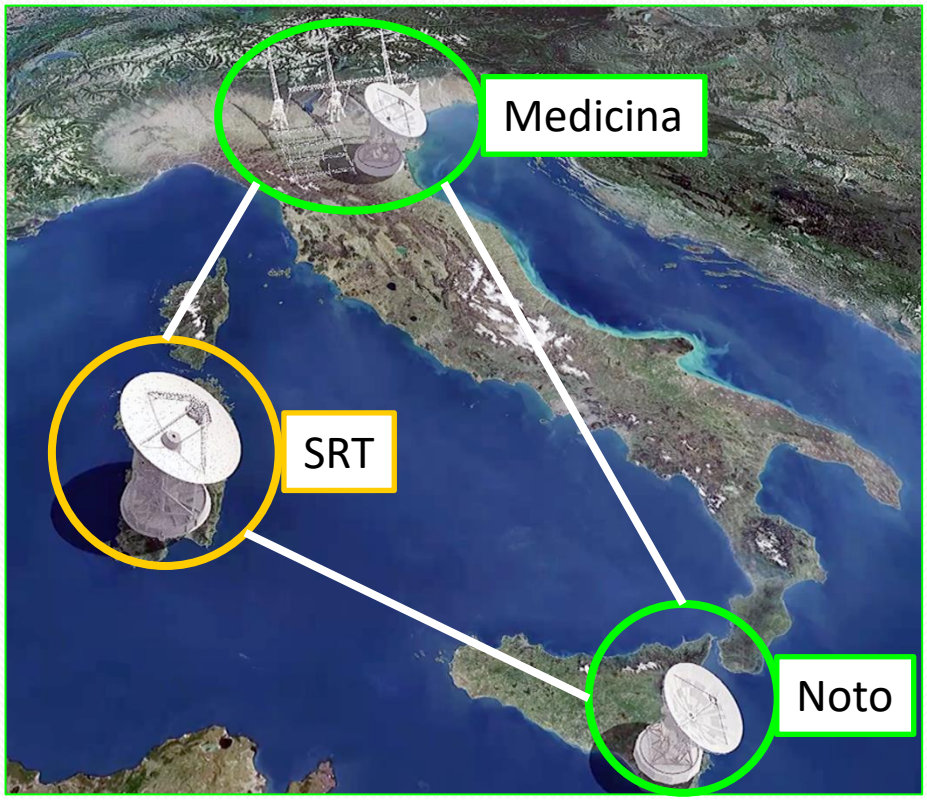
Trieste Solar Radio System 1.0 (now @IA2) and 2.0 (operational 2019 Q3)
(PI Mauro Messerotti)

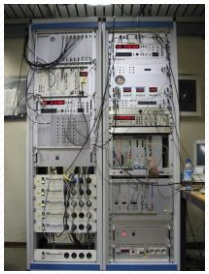
The Italian Radio Telescope Network

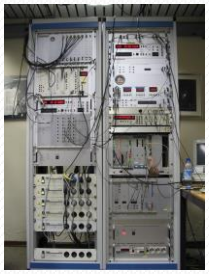
- INAF currently manages three fully-steerable radio telescopes

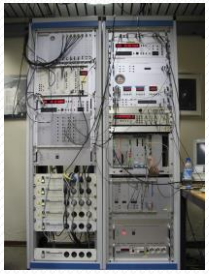
  (+ the Nothern Cross interferometer)

- Harmonization and coordination of efforts and resources at **INAF-IRA** and **INAF-OAC**




Single Dish

Very Long Baseline Interferometry

International VLBI Service for Geodesy

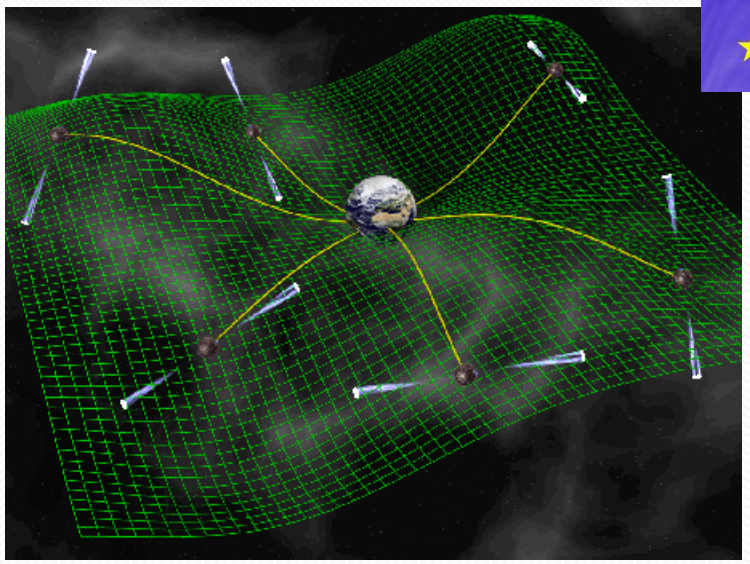
In Italy:

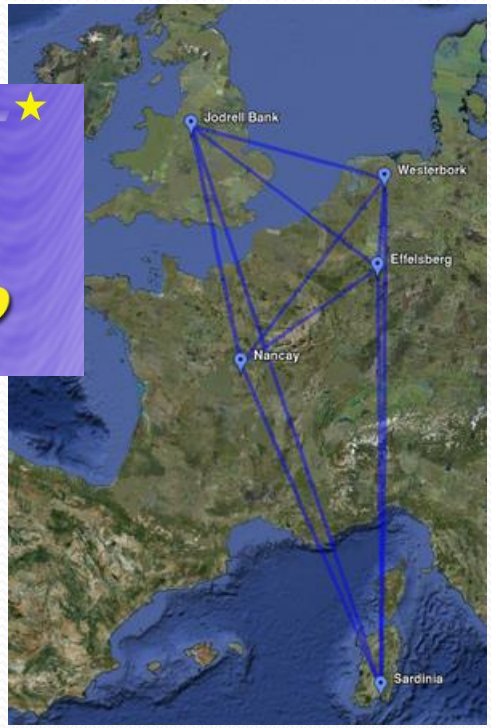
- Medicina and Noto (INAF)
- Matera (ASI)

Large European Array for Pulsars (LEAP)

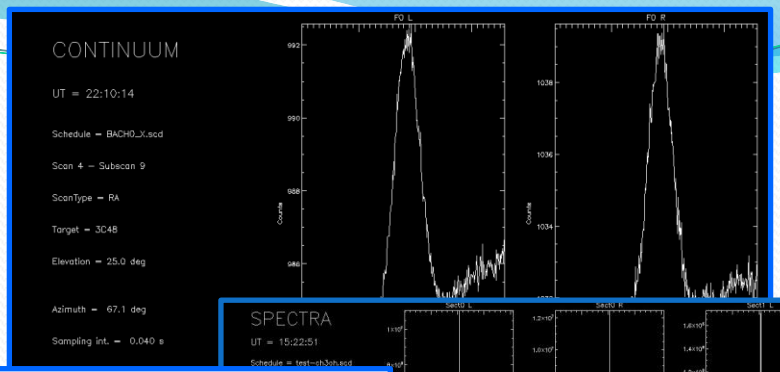
SRT
Sardinia Radio Telescope

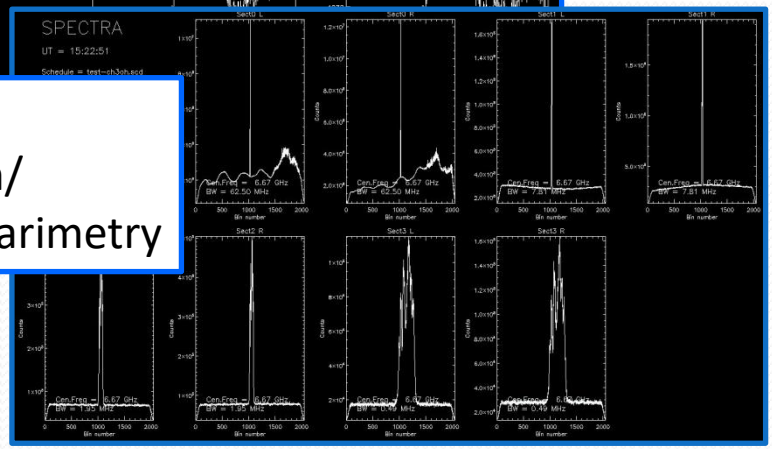
EPT
LEAP



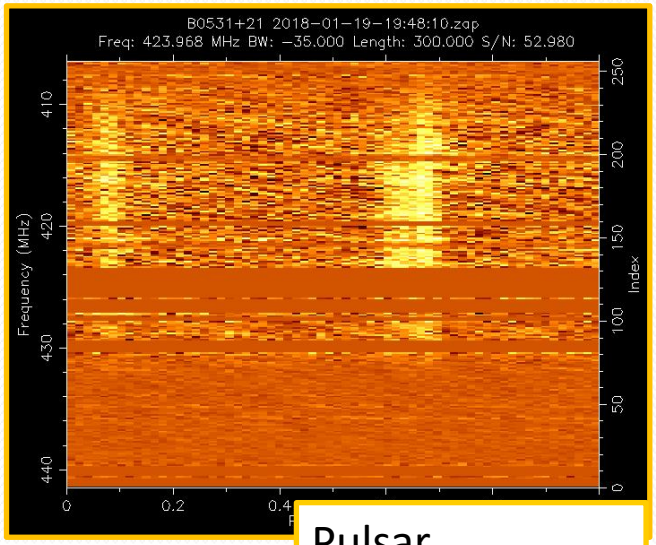


Many different data types needing
a common data archive

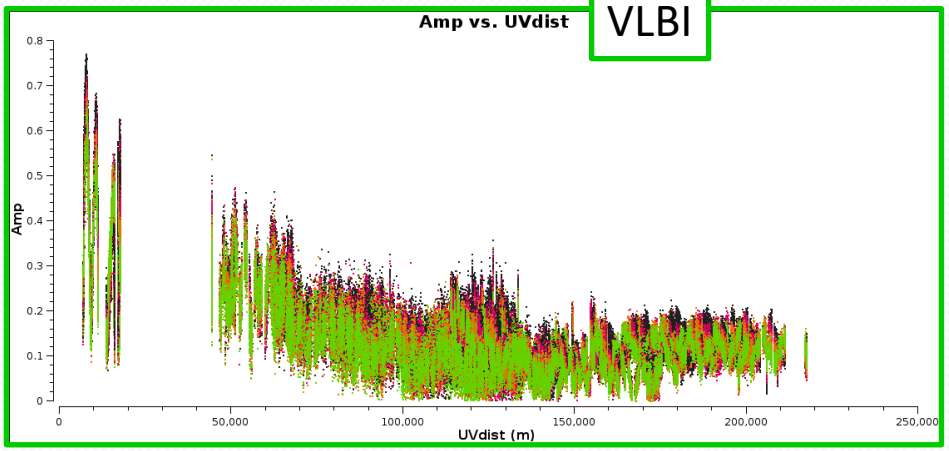




Single Dish
Continuum/
spectropolarimetry

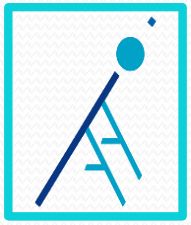


Pulsar
Single Dish/VLBI



VLBI

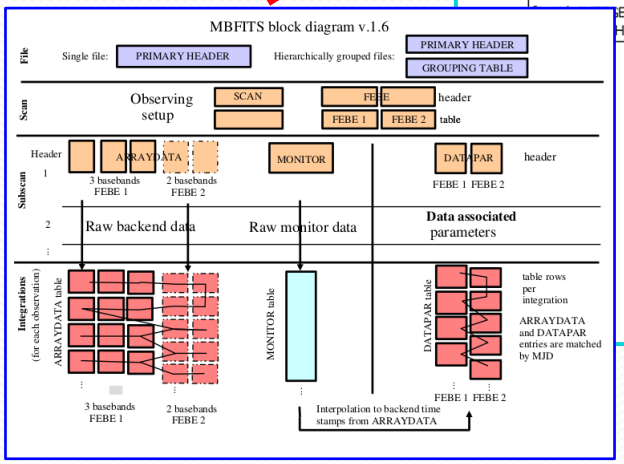
The INAF Radio Data Archive

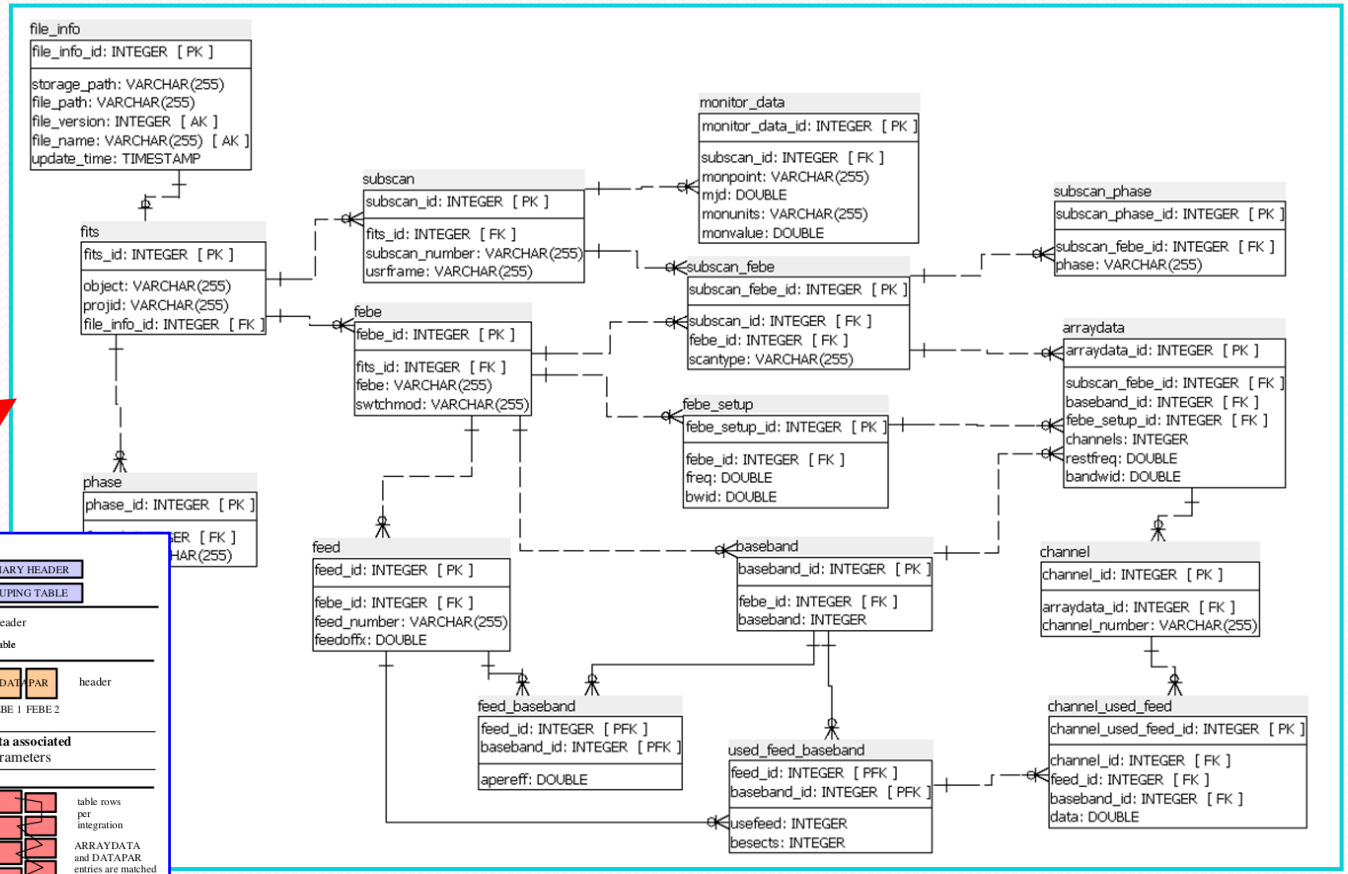


In collaboration with the INAF Astronomical Archives infrastructure

- Real-time data access for the (inter)national community
- Encourage scientific collaboration through metadata publication
- Data preservation

A data model for the archive

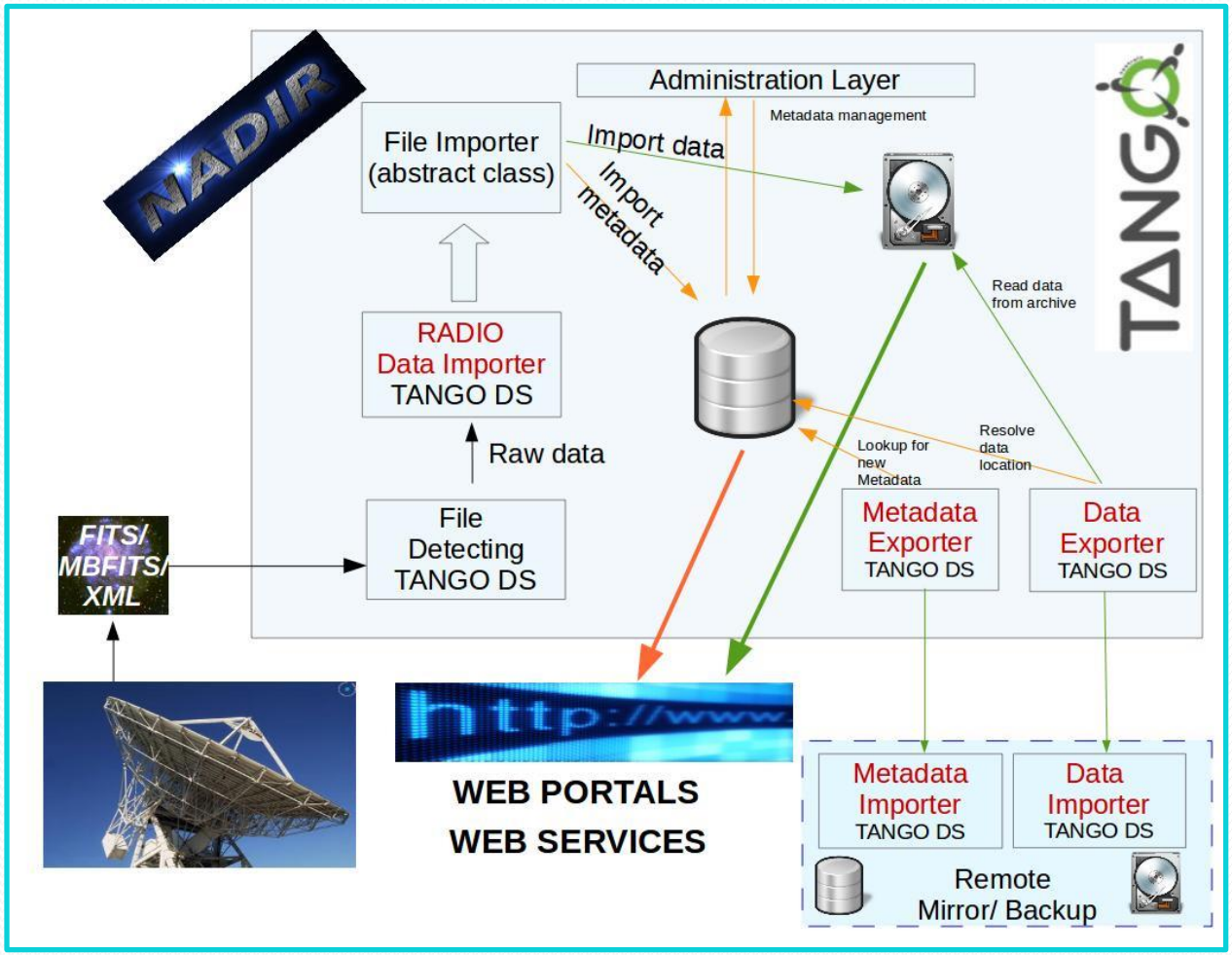




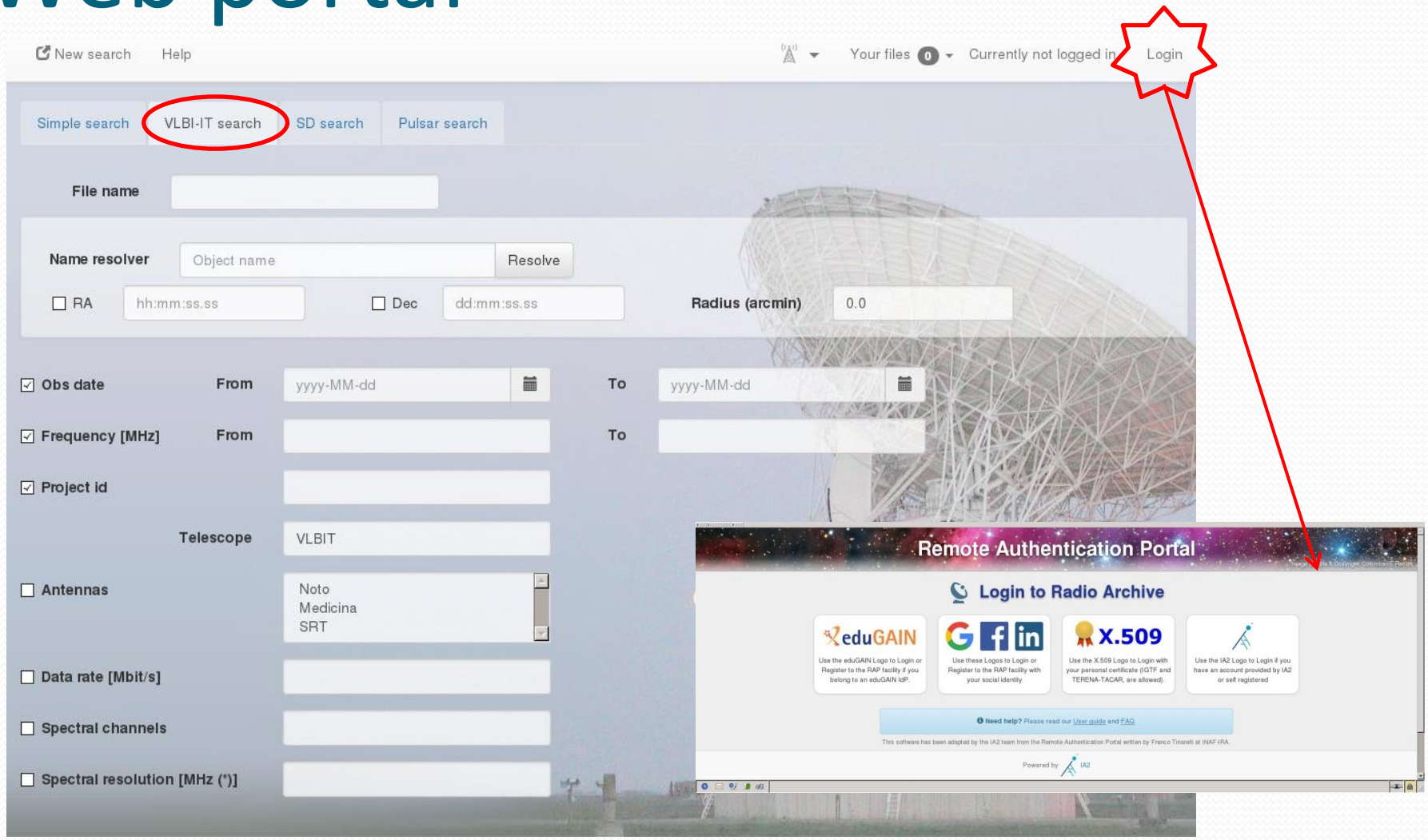
Data ingestion:
local @telescopes

Data storage: may
also be local

Common web portal
for all the telescopes/
observing modes



Web portal


New search
Help
Your files 0
Currently not logged in
Login
Simple search
VLBI-IT search
SD search
Pulsar search
File name
Name resolver
Object name
Resolve
RA
hh:mm:ss.ss
Dec
dd:mm:ss.ss
Radius (arcmin)
0.0
Obs date
From
yyyy-MM-dd
To
yyyy-MM-dd
Frequency [MHz]
From
To
Project id
Telescope
VLBIT
Antennas
Noto
Medicina
SRT
Data rate [Mbit/s]
Spectral channels
Spectral resolution [MHz (')]
Remote Authentication Portal
Login to Radio Archive
eduGAIN
X.509

Current challenges

To maximize the scientific return of the Italian radio telescopes the archive has to be easily exploitable also by a user with little or no expertise in radio astronomical data

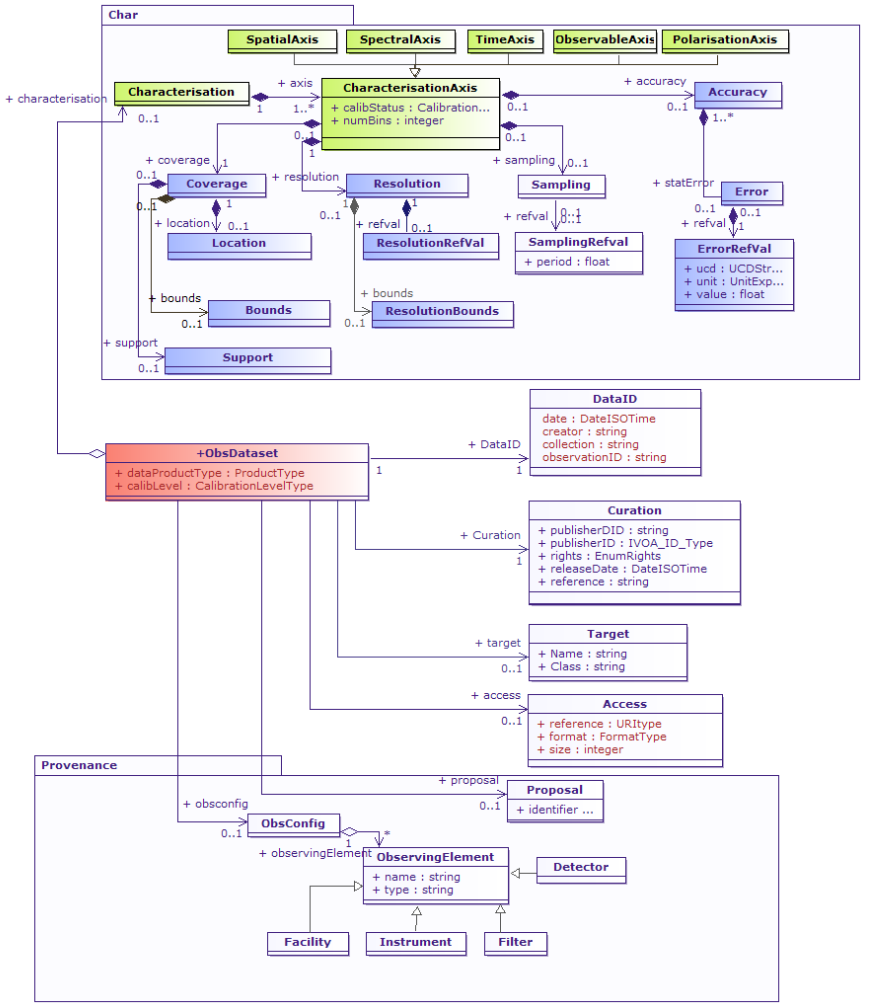
Archiving raw data that must be **discoverable** and **usable**…keeping in mind that also their size is becoming a concern…

Challenge n.1 - Discoverability

VO ObsCore

Global data discovery and access requires to expose
a uniform, standard data model

Verification of the VO ObsCore mandatory components for data discovery purposes

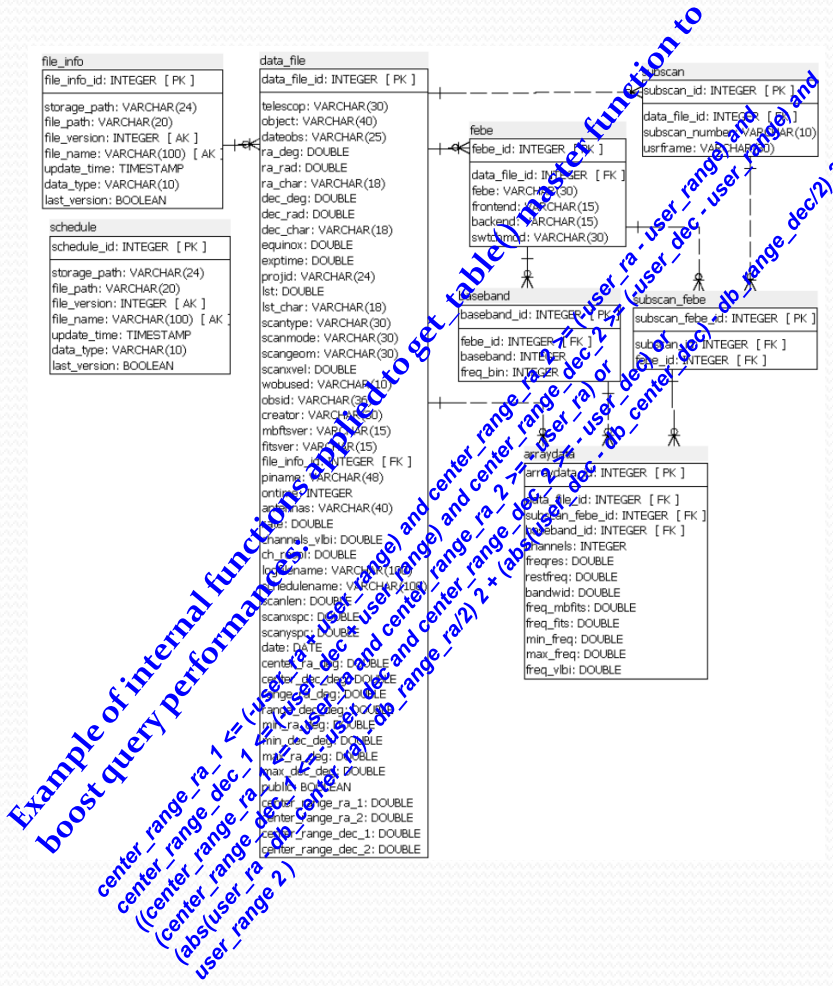


Challenge n.1 - Discoverability

TAP and DataLink

- TAP service:
  - IA2 implementation
  - custom User Defined Functions to improve query performance

- DataLink
  - Access to complex datasets: content list, ancillary resources (related datasets)
  - additional metadata (provenance, data quality, etc.)



Big Data

Handling big data: a concern not only for the new-generation interferometers!

The SARDARA spectrometer for Single Dish: real case 16000 spectral channels, with K-band 7-beam receiver, acquisition rate 33 ms gives an **effective data rate 160 GByte/hour**

Pulsar data: on average 150 Mbyte to 10 Gbyte per data file for folding mode acquisition, **from 3 to 400 Gbyte per data file** for search mode

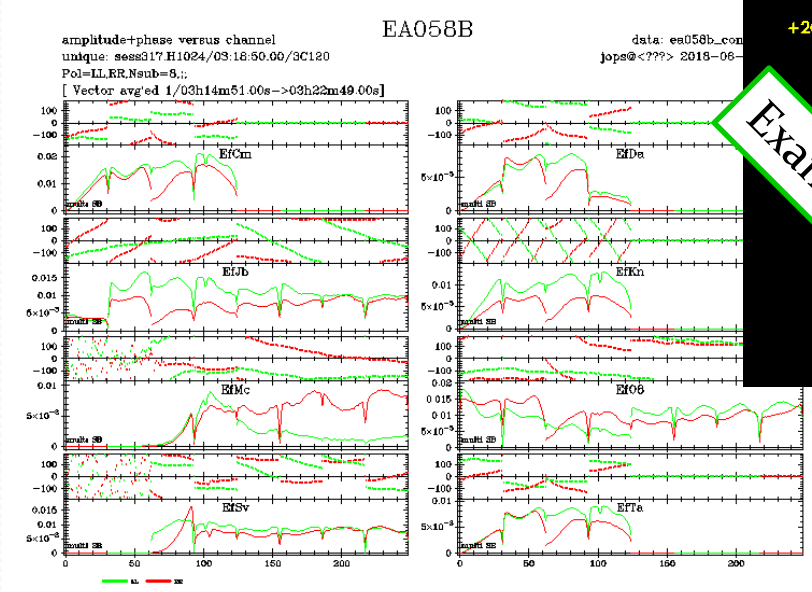
VLBI data: not yet a concern, but **strongly dependent** on the number of antennas, bandwidth etc.

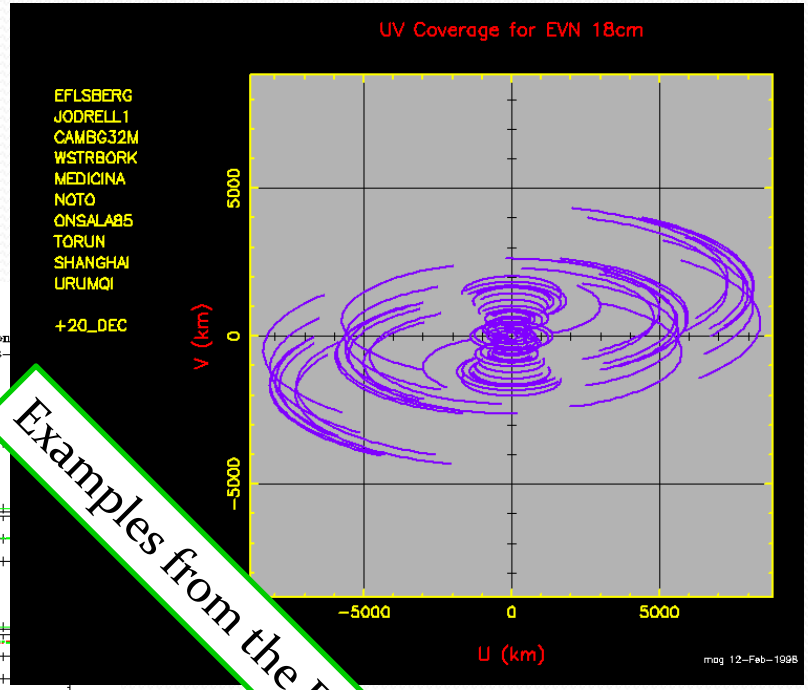
Challenge n.2 - Usability

Need to characterize raw data to help deciding if they are ok for the user's scientific purposes

Define effective metrics for data comparison and selection

Instrument characterisation

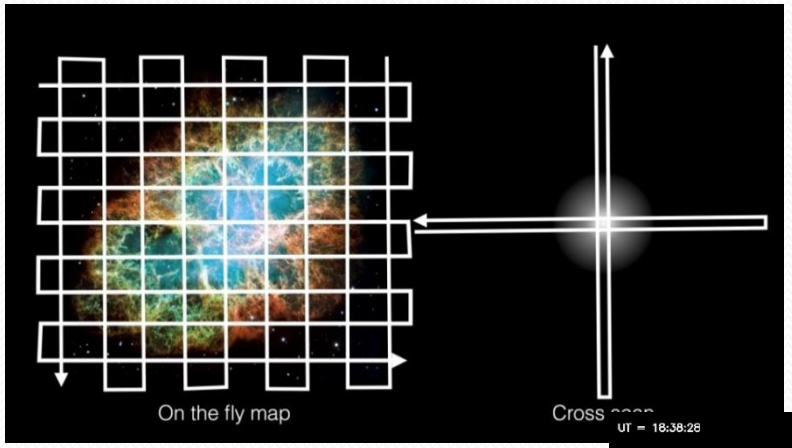




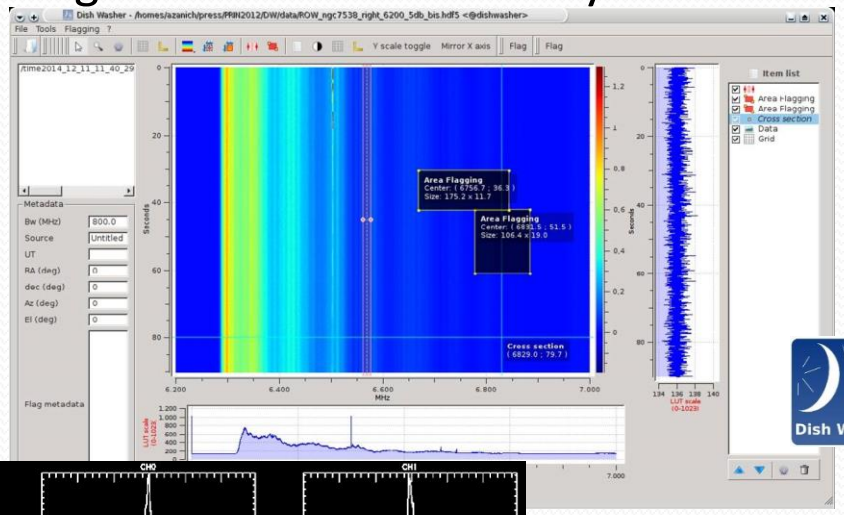
Examples from the EVN archive

Challenge n.3 - Software to the data

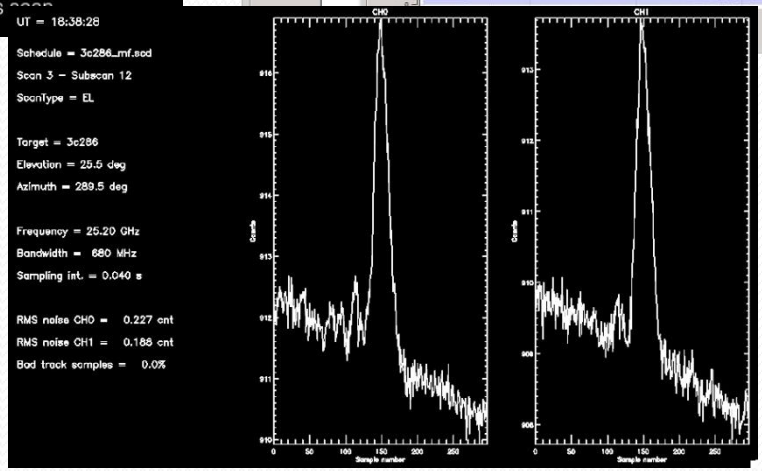
Data may be huge and/or their processing on local resources may be inefficient



SDT (Pellizzoni+)



DW (Zanichelli+)



CAP (Righini+)

Challenge n.3 - Software to the data

Provide user spaces: where, what, how?

Transform current tools into pipelines/create new software/analysis tools

Custom software

Standardisation

Data Flow

Quality Control

Archiving of data products (provenance etc.)

Conclusions

Radio data archiving in INAF is a reality, and more to come (LOFAR, SKA?,…)

The INAF radio data archive for the Italian radio telescopes

The radio telescopes archive is facing challenges that are somewhat similar to those of larger facilities

Towards an infrastructure for discoverability and usability of data

Big data and big challenges: characterisation of data to describe their scientific content. User spaces for data processing and analysis

The Italian radio telescopes archive seen as a possible test bed for future infrastructures?