The Italian radio data archive:

data content and provenance information

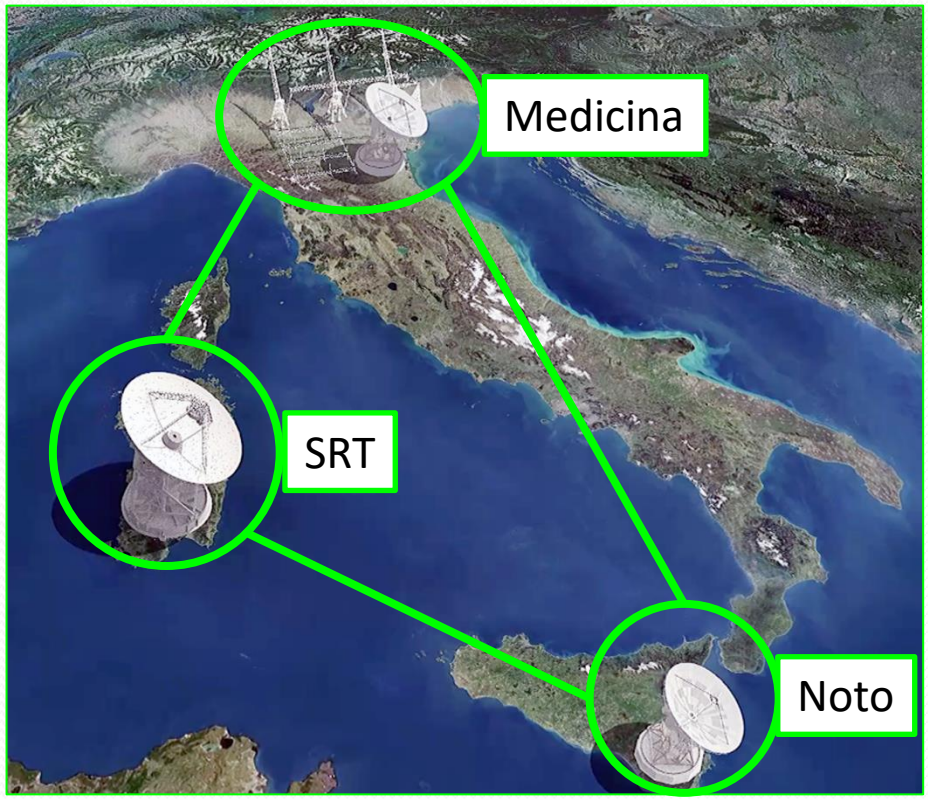
Alessandra Zanichelli

INAF – Istituto di Radioastronomia

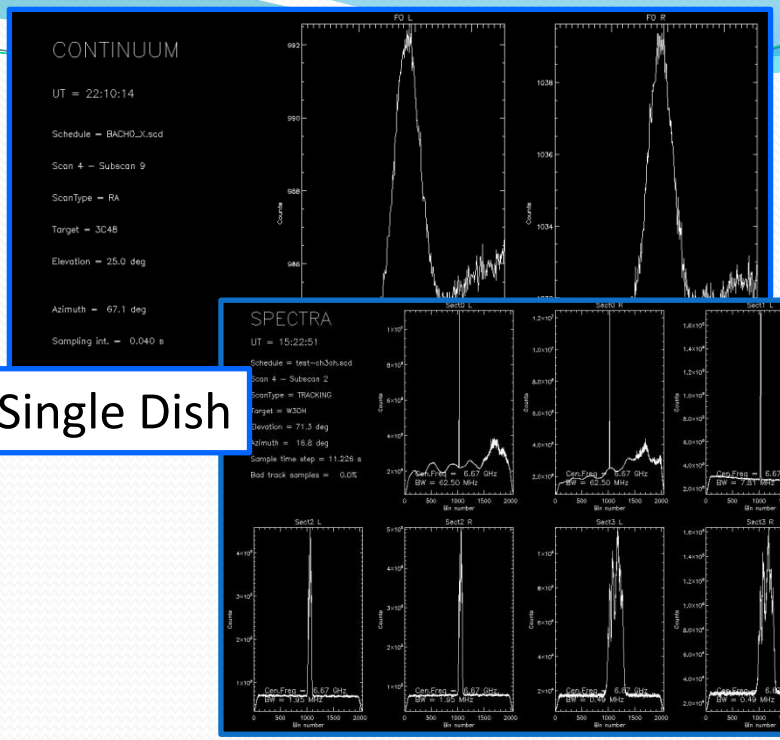
The INAF Radio Telescopes

Observing modes:

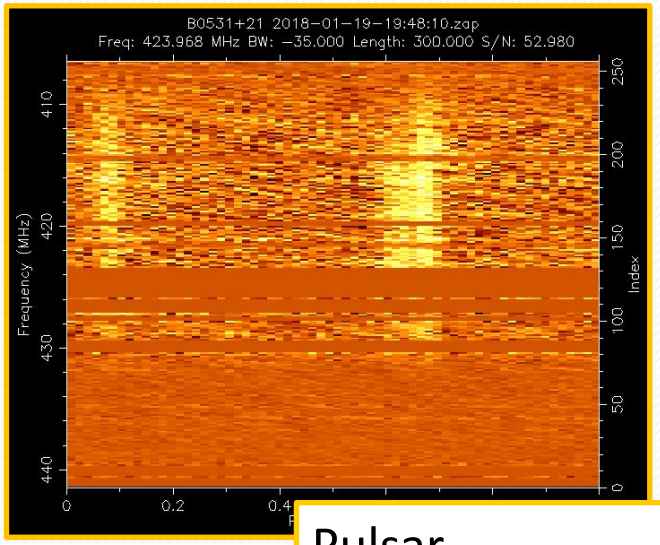
- Single-dish
- VLBI (EVN, IVS, EAVN)
- **VLBI-it**: *the «EVN lite» concept applied* (+software correlator)



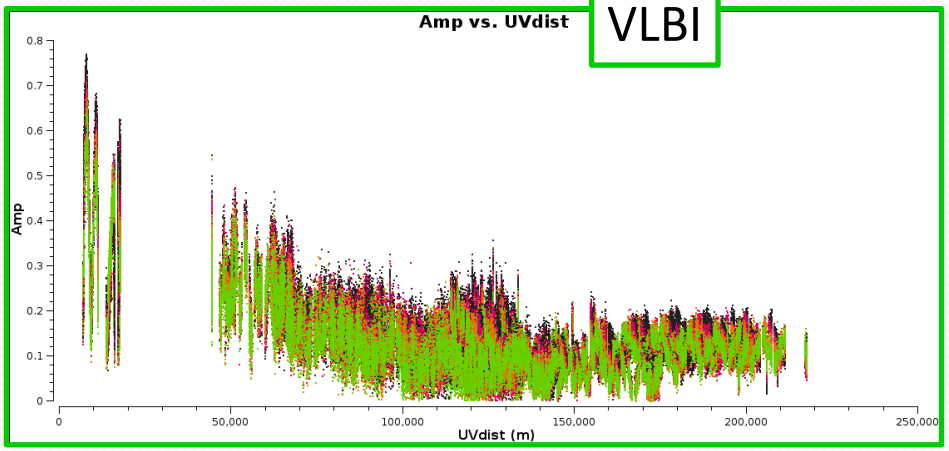
Different data types



Single Dish

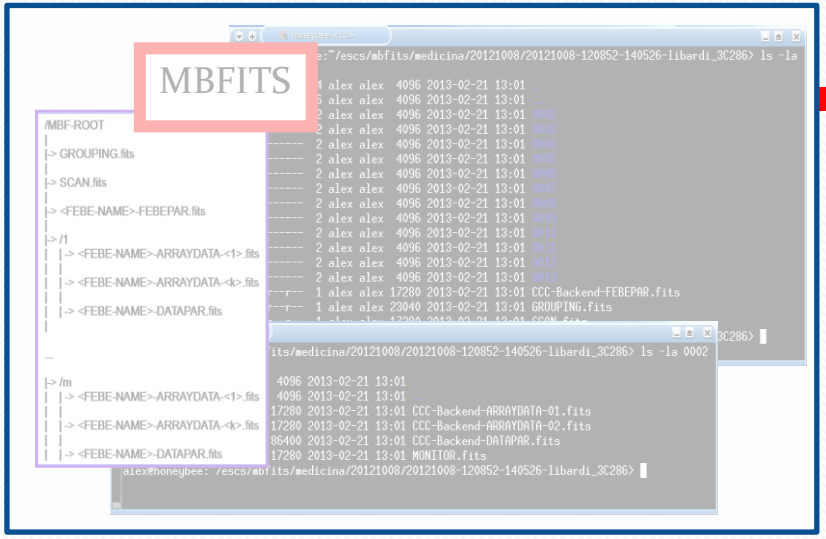


Pulsar
Single Dish/VLBI

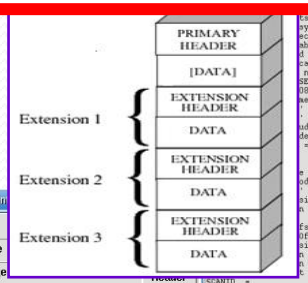


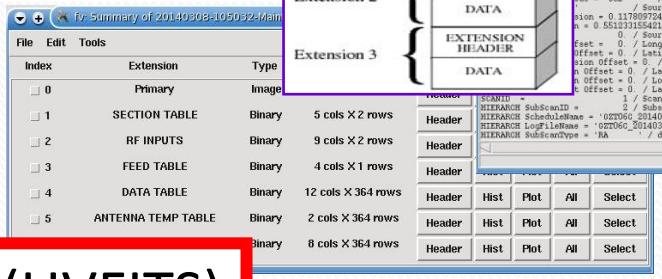
VLBI

Data formats

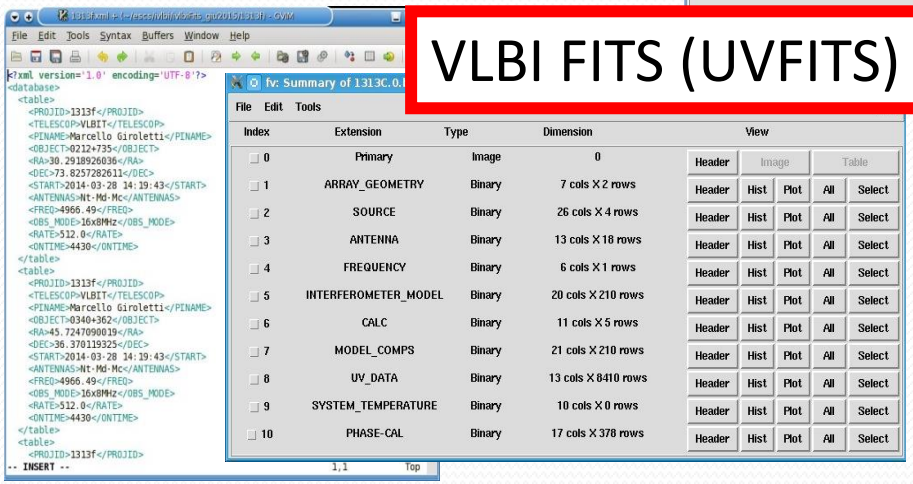


Single-dish FITS





VLBI FITS (UVFITS)



Pulsar data (PSRFITS)

The INAF Radio Data Archive

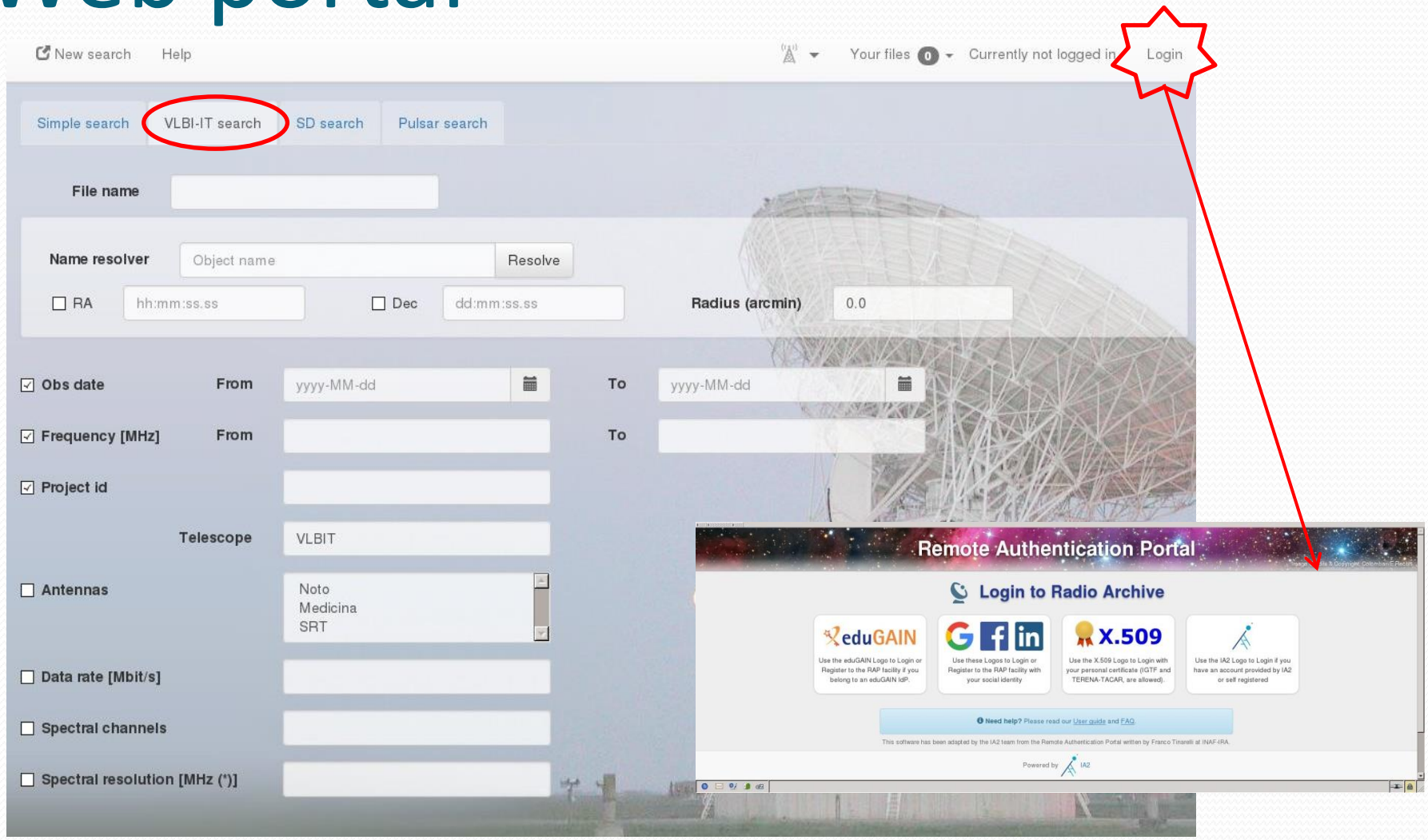
In collaboration with the INAF Astronomical Archives infrastructure

- Continuum and spectropolarimetric raw data from single-dish, pulsar and VLBI observations

- Archived:
  - ✓ The scientific exposure with [data+instrument+site] metadata
  - ✓ The observing schedule and logfiles

(Future: storage of processed data; Science Gateway and User Space)

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

Provenance

- Archived raw data must be discoverable and (re)usable

- Variety of observing projects and heterogeneity of the data ➡ accurate characterisation of the dataset is mandatory for the scientific exploitation of the Archive

- A «generic» Archive user (not the PI) must be able to address
  - ✓ if the data are suitable for her/his own research
  - ✓ if all the necessary information for data processing is available (e.g. calibration observations)

Provenance information

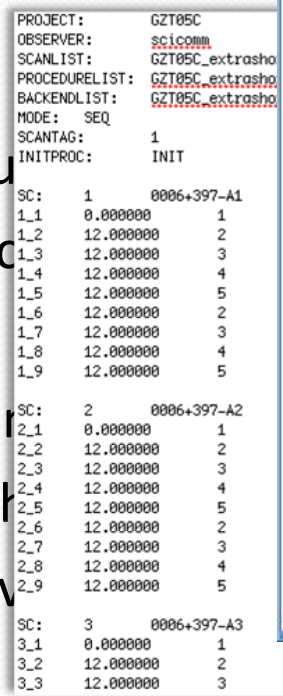
- Provenance DM not yet used but provenance information already stored for *raw data* and included in the Archive search parameters

- Metadata in the headers of archived files, for instance:
  - ✓ unique project ID, observer name, software version
  - ✓ telescope/instrument, etc.

- Ancillary information univocally associated to the dataset, for instance:
  - ✓ weather parameters
  - ✓ receiver performance, etc.

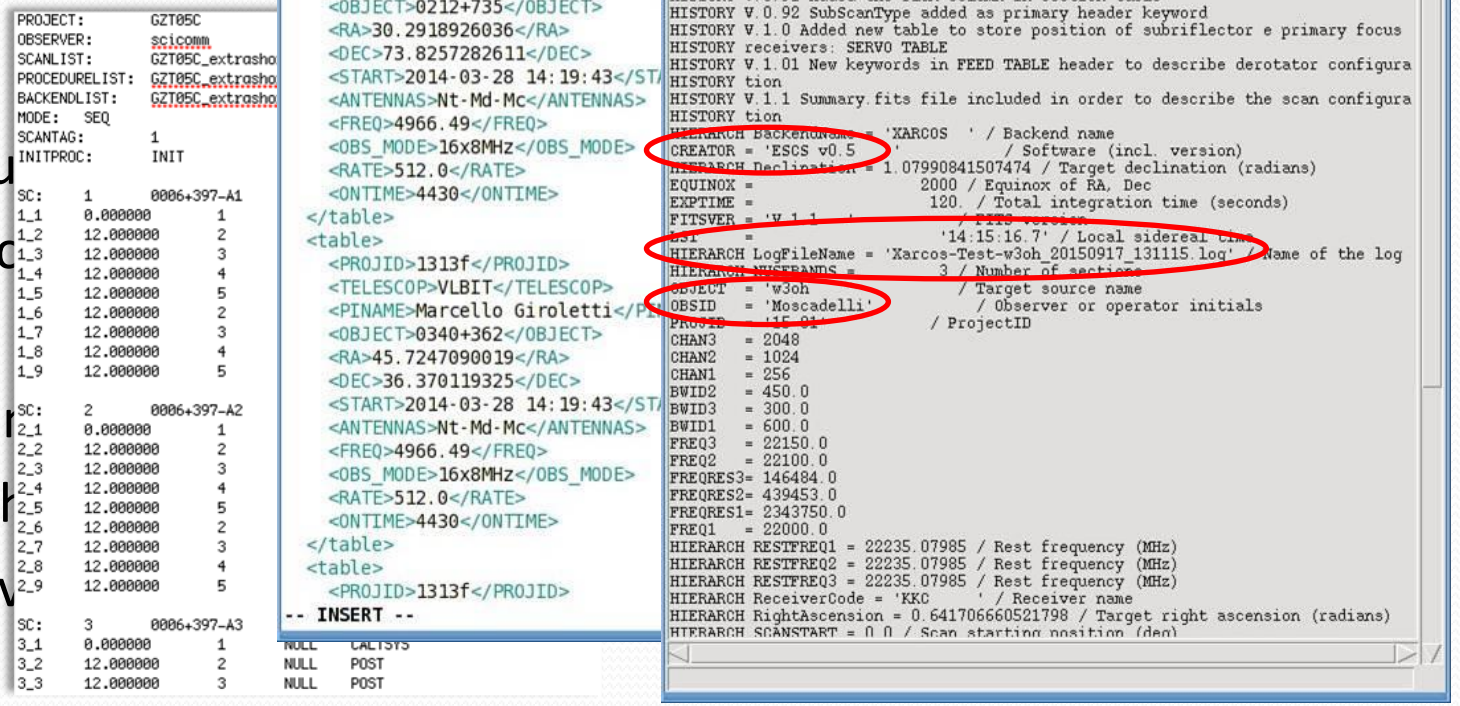
Provenance infor

- Provenance DM not
  stored for *raw data* a
- Metadata
  - ✓ uniqu
  - ✓ telesc
- Ancillary i
  - ✓ weath
  - ✓ receiv



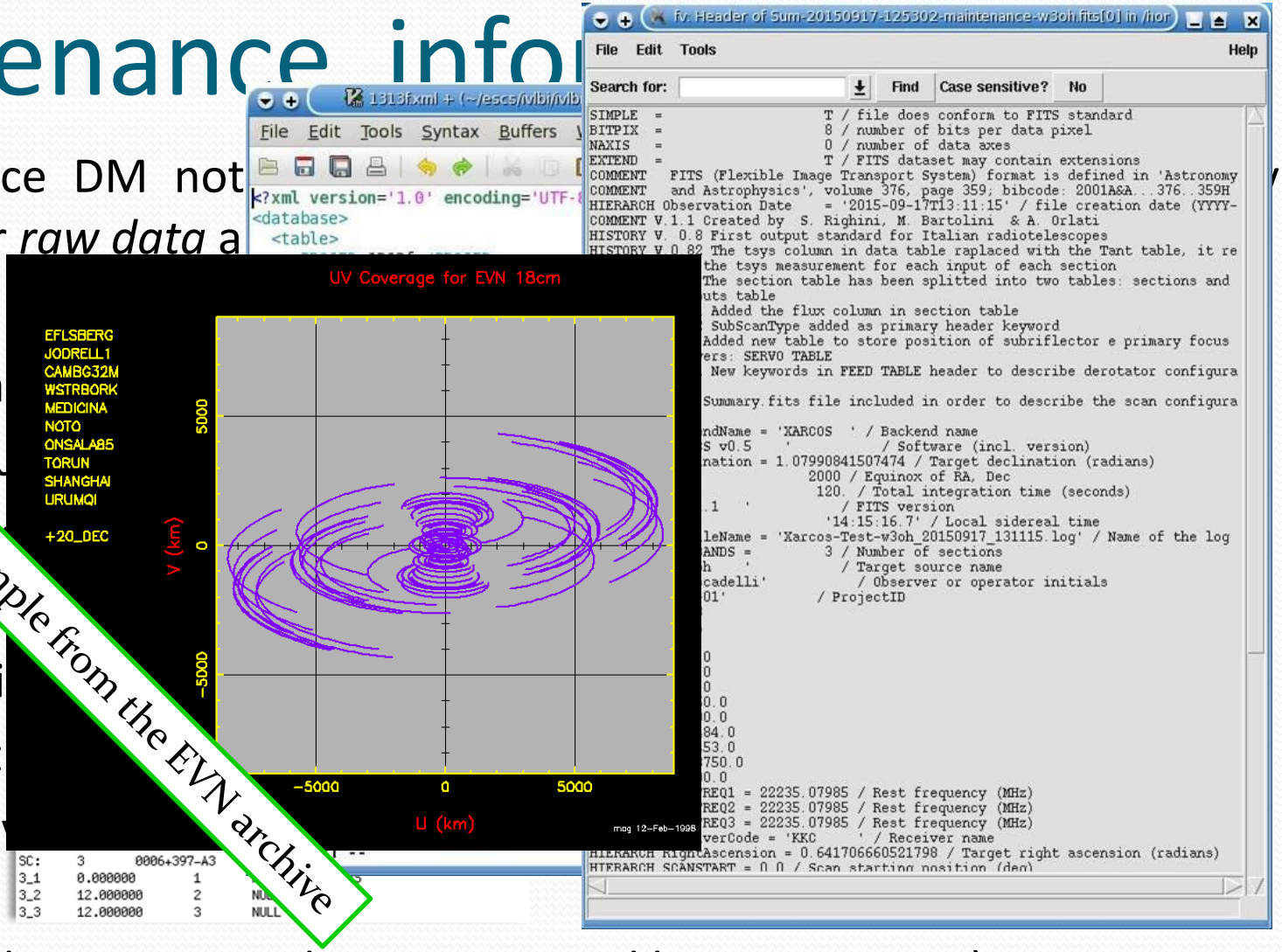
Provenance infor

- Provenance DM not
  stored for *raw data* a
- Metadata
  - ✓ uniqu
  - ✓ telesc
- Ancillary i
  - ✓ weath
  - ✓ receiv



Provenance infor

- Provenance DM not
  stored for *raw data* a
- Metadata
  - ✓ uniqu
  - ✓ t
- Ancillary i
  - ✓ weat
  - ✓ recei
- (Graphical: instrument characterisation like UV coverage)

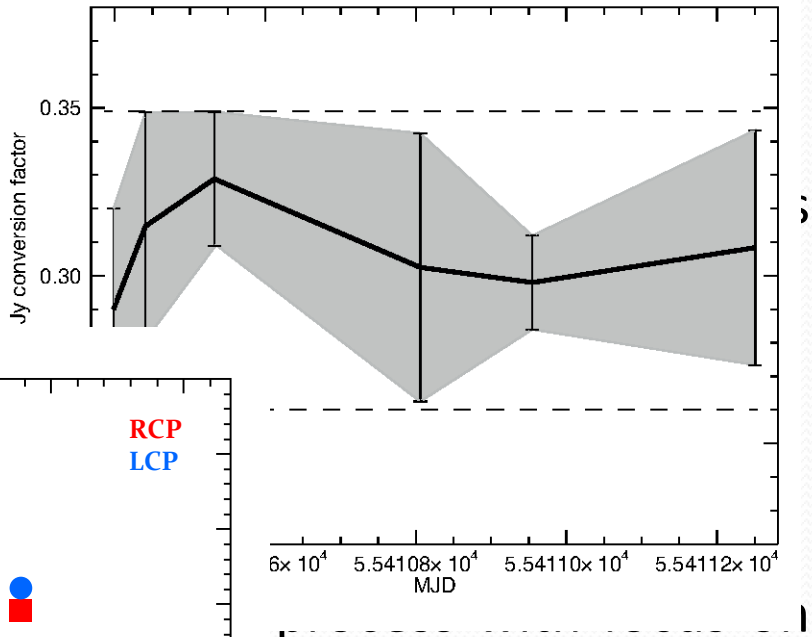


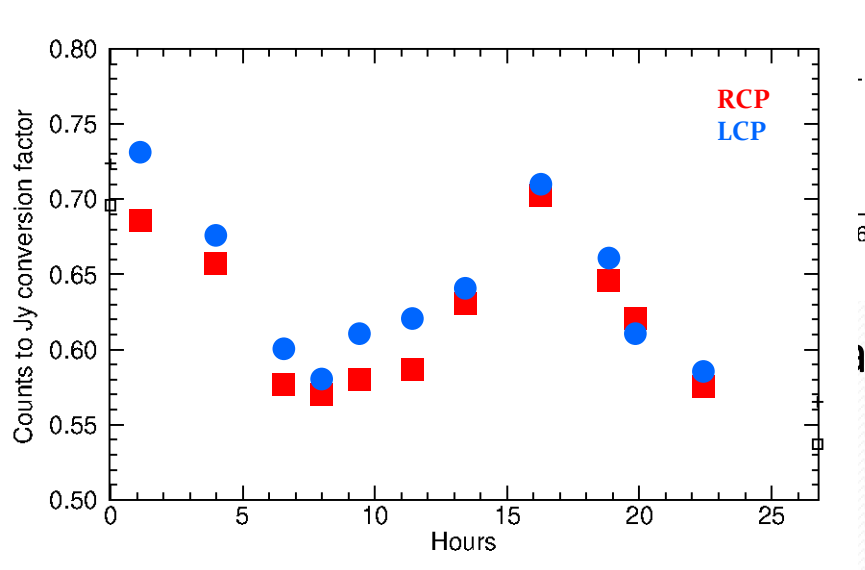
Provenance information (next)

- For *processed data*, provenance will include information necessary to fully describe the reduction history of the original dataset
- Processing pipelines, calibration and processing information as well as some level of quality metrics to be provided.
- Graphical information to visually describe specific characteristics
- Additional metadata to describe the reduction process with focus on the calibration steps (like RFI and atmospheric opacity removal)

Provenance information (next)

- For *processed data*, provenance will include information necessary to fully describe the reduction history o
- Processing pipelines, calibration and some level of quality metrics to be p
- Graphical int
- Additional n the calibrati acity removal)





Summary and future steps

- Different data types (and products) in the Archive
- Provenance information already present in raw data
- Raw datasets characterised by metadata and ancillary information
- *Archive processed data products and processing information. Effective quality metrics.*
- *IVOA Provenance DM for the interoperability of data.*
- *Level of granularity for provenance: critical for reproducibility*