A Public Data Archive for the Italian Radio Telescopes

A. Zanichelli[a], C. Knapic[b], S. Righini[a],
M. Nanni[a], M. Stagni[a], F. Bedosti[a],
M. Bartolini[a], M. Sponza[b], A. Orlati[a],
R. Smareglia[b]

a – INAF-Istituto di Radioastronomia
b – INAF-Osservatorio Astronomico di Trieste

The amount of data delivered by modern instrumentation and observing techniques is bringing radio astronomy in the era of Big Data, and the nowadays widely adopted Open Data policies allow free and open access to data from radio astronomy facilities. A fundamental ingredient to enable Open Science in the radio astronomical community and to engage also public participation (the so called Citizen Science) is thus the availability of public archives in which data can be accesses and searched with modern software tools.

A web-based, VO-compliant public Archive has been built to host data from the Italian radio telescopes managed by the National Institute for Astrophysics (INAF). The archive main features consist in the capability to handle the various types of data coming from the different observing instrumentation at the telescopes; the adoption of a policy to guarantee the data proprietary period; the accessibility of data through a web interface and the adoption of VO standards to allow for successful scientific exploitation of the archive itself in the data mining era.

The Radio Data Model

The three Italian radio telescopes at Medicina, Noto and Pranu Sanguni can be used separately as single-dish instruments (SD) or in a coordinated manner within the (inter)national interferometric network (VLBI-IT). The two observing modes are characterized by different output (meta)data formats, typically FITS, MBFITS and XML. The definition of a common **Radio Data Model** has been the first step towards a common archive. The Radio Data Model was built on top of the data/metadata structure defined in the MBFITS standard (APEX-MPI-ICD-0002 by Dirk Muders), produced for the Atacama Pathfinder Experiment (APEX). Our model is capable of handling radio data written in non-hierarchical FITS format as well, in order to serve a vaster range of users.

Data Preparation and the Radio Data Importer

Given the different data formats in use and in particular the hierarchical structure of SD datasets, a dedicated software (the *Finalizer*) has been developed for data preparation before archive ingestion. Aiming at persistence and future exploitation of data, the ancillary information contained in the observing schedules and telescopes/correlator logs is archived for each dataset as well.
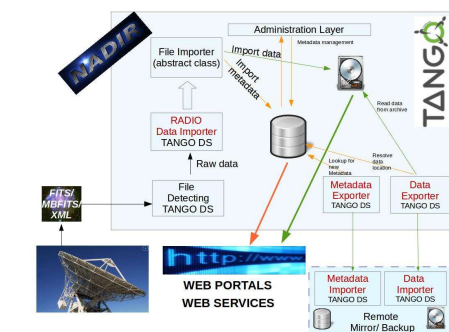
The Radio Data Importer (RDI) is a TANGO1 server for importing in the Radio Archive database the FITS, MBFITS and XML files containing the observational metadata. RDI configuration is stored in the datamodel database, which determines for each observing mode the set and the structure of data that has to be read from the input files, as well as the storage directory. A general MBFITS database is used as a baseline for the creation of the Radio Archive database, which actually stores the data retrieved from the input files according to the RDI configuration. Such a structure allows for easy integration of new and different instruments in the Radio Archive, provided their data can be described by means of the general MBFITS database.

The Finalizer, RDI and the data storage are geographically distributed and run independently at the various telescopes.

The Radio Archive Architecture

- Open source
- Based on TANGO Distributed Control System (independent from OS)
- Modular, flexible configuration
- Scalable in n. of RDI devices (data ingestion is locally distributed at the telescopes)
- Handling of different data format (FITS/ MBFITS/ HDF5/ XML)
- Flexible revision of policy and versions
- Development of services in the major OO Languages (C++, Java, Python)
- Strong logging and error handling
- Maintained at INAF-Italian Astronomical Archives Service (IAA)

The Radio Archive on the Web

Access to the Radio Archive is possible through a dedicated web interface. Possible archive users in the Open Science era include scientists and telescope staff but, in principle, also the general public. To guarantee the data proprietary period according to INAF rules, a Single-Sign-On login authentication mechanism is foreseen.

Users perform dedicated queries in the Radio Archive by means of web forms. Query parameters can be generic, common to SD and VLBI-IT like for instance the celestial object coordinates, or specific to the observing mode, like the scan geometry or the subset of antennas. The output page contains a summary of query results in tabular form, to allow the download of checkbox-selected datasets and optionally also ancillary information (schedules and log files).

To improve the query performance and speed of results delivery, an indexing process of the most commonly used columns and SQL functions is implemented. To limit real-time calculation, all the parameter computations that could be done during the query execution are performed in advance and stored in the database following the OAIS (Open Archival Information System) standard rules. Also, indexes were created on various columns and combinations of columns. These processes are customized on the necessities of a Radio Archive query form, and on the effective necessities of the average radio astronomer curiosity and needs.

The Radio Archive and the Virtual Observatory

In view of improving open data re-usage and the accessibility of scientific data to the astronomical community and the general public, VO-compatible standards like the Table Access Protocol (TAP) are preferable. To this aim, a TAP-based service for the Radio Archive Database is going to be published. The use of Observation Data Model Core Components and its appropriate application to the radio astronomical raw data from the Italian Radio Telescopes is also under study (see Cristina Knapic's talk) to investigate how and to what extent the VO already addresses the use cases for the Italian Radio Archive.
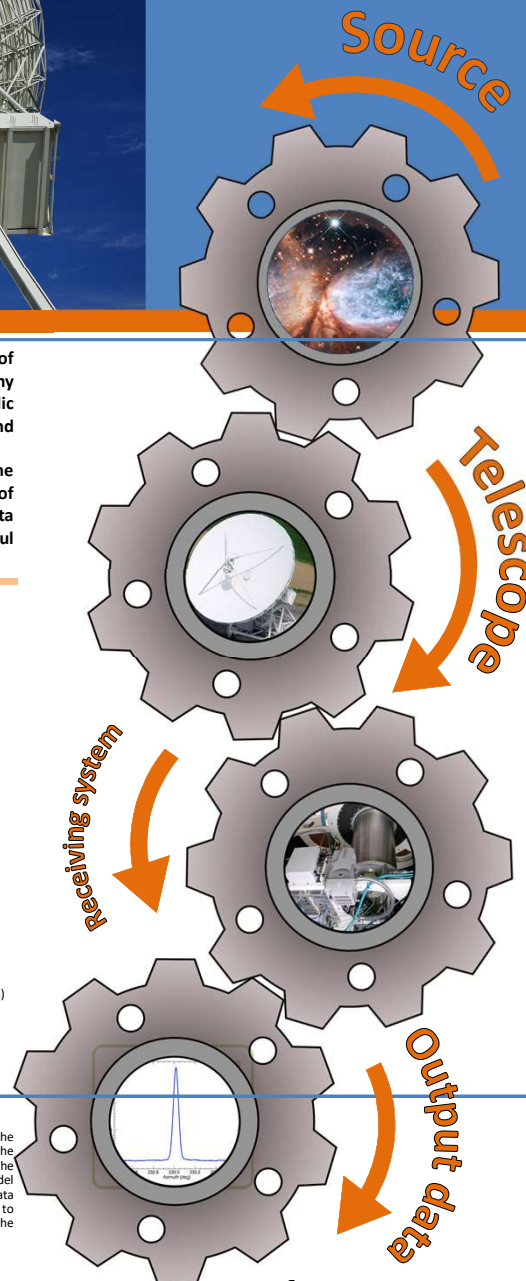
For details on our antennas and all the related facilities visit our webpage devoted to the observers.

Source

Telescope

Receiving system

Output data

Data formats

Radio Archive

ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS