Data formats for the Medicina and Noto radio telescopes in view of a common Archive

A. Zanichelli[1], S. Righini[1], C. Knapic[2], M. Stagni[1], M. Nanni[1],
M. De Marco[2], E. Dovgan[2], A. Orlati[1], F. Bedosti[1], R. Smareglia[2]

Version 5.7 – January 2016

INAF ORA-BO
Report no. 488/15

Referee: F. Tinarelli[1]

[1] INAF ORA-BO, via Gobetti 101, I-40134 Bologna, Italy
[2] INAF Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34131 Trieste, Italy

CHANGE RECORDS

Issue	Rev.	Sections	Page	Date	Comments
1	0	All	All	Nov 2013	Created
1	1	All	All	Nov 2013	Revised/completed
2	0	All	All	Nov 2013	Modified format for SD data, and relative sections
3	0	4, 5		March 2014	Revised.
4	0	4, 5, 6, Appendix A	All	March 2014	Added Section 6. Revised the others.
5	0	4, 4.1,5,6, Appendix A	All	April 2014	Change in the structure of FITS dataset. Added Appendix for Summary.fits file description
5	1	All	All	April 2014	Revised.
5	2	Sects. 3, 4, 5, 6	All	Oct 2014	Schedule+logfile coupling with SD data changed. Some SD keywords defined. Format of VLBI-IT data parameters file defined. New section 6
5	3	Sect. 9	All	Dec 2014	Added section 9 with description of the archive web interface added
5	4	All	All	April 2015	Global revision.
5	5	All	All	May 2015	Refereed version, published in IRA Report Library.
5	6	All	All	Dec 2015	Modified some keyword names/formats. Added multiplicity for FREQRES, RESTFREQ. Hopefully final syntax definition for some keywords. Redefined the syntax of SD summary file name. Clarified the selection of antennas in VLBI-IT queries.
5	7	All	All	Jan 2016	Updated the list of VLBI-IT query parameters, clarified some SD parameters.

Table of Contents

LIST OF ACRONYMS

EVN	European VLBI Network
FITS	Flexible Image Transport System
HDU	Header/Data Unit
INAF	Istituto Nazionale di Astrofisica / National Astrophysics Institute
MBFITS	Multi-Beam Flexible Image Transport System
PI	Principal Investigator
SD	Single Dish
SRT	Sardinia Radio Telescope
VEX	VLBI Experiment
VLBI	Very Long Baseline Interferometry
VLBI-IT	Italian VLBI Interferometer
XML	eXtensible Markup Language

1 INTRODUCTION

This document is intended to give an overview of the various formats for data acquired with the Medicina and Noto radio telescopes that will be stored and handled by a common radio Archive. A description of the Archive architecture is presented as well. This work is a collaboration between the Istituto di Radioastronomia and the IA2 infrastructure at the Osservatorio Astronomico di Trieste and aims at providing the (inter)national community with a state-of-the-art archive for radio astronomical data. The design of database structure and data formats is well in place, as described in this paper, and a first Archive prototype is up and running as of May 2015. Some minor adjustment may be required and will be detailed in future revisions of this document.

2 GENERAL FEATURES

The IRA radio telescopes at Medicina and Noto can observe in Single-Dish (SD) or VLBI mode. Up to now, observers were in charge of saving data coming from their SD observations, while data from VLBI observations performed by Medicina and/or Noto within the European VLBI Network (EVN) are usually stored in the EVN Database.

Recently, it has been decided to equip the IRA radio telescopes with an Archive for the systematic storage of the SD observations. Moreover, with the advent of the new Sardinia Radio Telescope it will be possible to realize a fully Italian VLBI array (hereafter VLBI-IT) by using the three antennas in a coordinated manner and applying software correlation for the pre-processing phase. Data coming from such pre-processed VLBI-IT observations will be stored in the same database used for SD data. The Archive must thus be able to host and handle two kinds of data: those coming from VLBI-IT observations and those coming from SD ones, with a variety of formats that will be described in the following sections.

The Archive should be queried through web-based interfaces by three main classes of users: the administrator for maintenance and/or statistical purposes; the project PIs, to retrieve their data during the proprietary period; the scientific community at large, to retrieve any of the publicly available data.

3 VLBI-IT DATA FORMAT

Each VLBI-IT observation is executed by a network of antennas including at least one of the three Italian radio telescopes plus telescopes from other countries. Each antenna is identified by an unique code, namely 'Mc', 'Nt' and 'Sr' for Medicina, Noto and SRT respectively. The size of the output data set *for each telescope* is of the order of 500GBy/hour.

The observing procedure follows this general scheme:

- The PI of the project prepares the observing schedule using the Sched (or Sked for geodetic experiments) software tool.
- Sched produces the so-called VEX file, which is a text file containing all the information that are relevant for the observation. An example of VEX file can be found at this address: http://www.ira.inaf.it/vlb_arc/vlb_archive/aug13/rp022.vex .
- The PI of the observational project uploads the VEX file in a common ftp site for VLBI experiments.
- Antenna operators download the VEX file and run it at the proper date/time to execute the observation.
- Each antenna operator uploads in the ftp site the antenna log file for the observation. Antenna raw data are transferred to the correlator site.

- Correlator operators download from the ftp site the VEX file and the log files of all the antennas involved in the experiment. Together with the data sets, these files are used to execute the correlation procedure.

The correlation produces a so-called Visibility Data file: this is a monolithic FITS file some Gby in size. This file contains the real raw-data for an interferometric experiment and is given in input to the data reduction tools for calibration, filtering and image reconstruction.

The Visibility File may contain data for more than one source: it's the summary of all the observations done for a given observing session. It may contain, for instance, data for a number of target sources plus data from observations of calibrator sources.

The Visibility File must be viewed as an "atomic unit" and as such will be archived. It is not recommended to split it into single-source with the exception of those cases in which the file size is too large. In fact, data inside a Visibility File are homogeneous and contain all the calibration information essential for subsequent data processing. Moreover, the Visibility File is the "unit" that will be delivered to the PI.

Being a FITS-like file, the Visibility File contains some header keywords needed by the most commonly used packages for data reduction like AIPS, CASA and MIRIAD. The keyword set is by no means complete in terms of the most commonly used query parameters (for instance, RA and Dec information are not written explicitly in keywords), and other relevant search parameters like field size or resolution are set only after the data reduction process is executed. Moreover the FITS file structure may vary in number and position of the HDU depending on the type of observations: imaging, satellite tracking, pulsar observations, etc. At the same time, the VEX file associated to the observation contains all the information that is relevant for the archive search, but given its structure it is not easy to be parsed in order to fill the database tables with all the relevant metadata. For these reasons it has been agreed that each FITS Visibility Data file will be accompanied with a Summary file (similar to the one used for single-dish FITS scans, see Section 4) written in XML format.

The VLBI-IT Summary.xml file will contain all the relevant information organized in blocks of keynames/keyvalues, *one block for each source* contained in the FITS Visibility Data file. In Appendix A an example of Summary.xml file is reported. With such a structure there is some level of redundancy in the parameters (e.g. The TELESCOP keyword is repeated for each source), nevertheless this assures a complete description of all the sources in a dataset.

Keyword names in the Summary file are kept as uniform as possible with those used for single-dish FITS/MBFITS data. In particular, it has been agreed that in the Summary.xml file the TELESCOP keyword will take the value 'VLBIT', to make VLBI-IT a distinct instrument with respect to other observing modes.

Following the prescriptions for SD data, we adopt also for the VLBI-IT summary file the naming convention: Sum-DATE-UT-PROJID-OBJECT.xml (see Sect. 4.1). In the following we will generally refer to this file as Summary.xml

For VLBI-IT data, it is foreseen the production of a comment file at the end of the correlation process listing e.g. the correlation quality, possible problems with some antennas and any other information that can be useful for subsequent data processing. This file is sent to the archive as well.

4 SD DATA FORMATS

Currently, SD data for the Italian radio telescopes are written in single-dish FITS format, while in the near future the MBFITS format will be applied. The former is a kind of table-based FITS which can be easily handled by the most common data analysis tools. The latter is rapidly becoming a standard in many of the modern radio and mm telescopes due to its ability to handle various types of instrumental configurations.

4.1 FITS

Single Dish data at Medicina, Noto and SRT can be written in FITS format as a series of binary tables according to the structure described in Table 1. With respect to prior versions of this document the extension

SERVO TABLE is now present in all the FITS data coming from SD observations with any of the Italian radio telescopes.

Table 1: HDU structure of a Medicina single-dish FITS file.

Index	Extension	Type
0	Primary	Image
1	SECTIONTABLE	Binary
2	RFINPUTS	Binary
3	FEEDTABLE	Binary
4	DATATABLE	Binary
5	ANTENNATEMPTABLE	Binary
6	SERVOTABLE	Binary

The various HDUs contain information needed to describe the observation and the instrument setup; f.i. in the case of a multifeed receiver the FEED TABLE describes the feeds, their relative position and power.

In the future this format will be gradually dismissed in favor of the hierarchical MBFITS one, but still the archive will have to host also this kind of data.

Each observation of a given source performed with a given instrumental setup (a "scan") may be composed of one or more FITS files (one for each "subscan"). All the FITS files belonging to the same scan are written inside a common folder (hereafter s*can folder*) together with a *Summary file* in standard FITS format, see below.

Single-dish FITS scan folder names are composed as: DATE-UT-PROJID-OBJECT, while the single FITS file naming convention is: DATE-UT-PROJID-OBJECT-SCAN-SUBSCAN. As an example here is the listing of the directory 20131210-202656-maintenance-3C123, which contains one observation (scan no. 7) made of eight subscans on the source 3C123:

```
/home2/escs/archivioDati/esempio_struttura_fits> ls -la 20131210-202656-maintenance-3C123
total 684K
-rw-r----- 1 azanich users 86400 Dec 12 10:20 20131210-202656-maintenance-3C123_007_001.fits
-rw-r----- 1 azanich users 80640 Dec 12 10:20 20131210-202715-maintenance-3C123_007_002.fits
-rw-r----- 1 azanich users 86400 Dec 12 10:20 20131210-202735-maintenance-3C123_007_003.fits
-rw-r----- 1 azanich users 86400 Dec 12 10:20 20131210-202755-maintenance-3C123_007_004.fits
-rw-r----- 1 azanich users 80640 Dec 12 10:20 20131210-202814-maintenance-3C123_007_005.fits
-rw-r----- 1 azanich users 80640 Dec 12 10:20 20131210-202834-maintenance-3C123_007_006.fits
-rw-r----- 1 azanich users 86400 Dec 12 10:20 20131210-202854-maintenance-3C123_007_007.fits
-rw-r----- 1 azanich users 80640 Dec 12 10:20 20131210-202913-maintenance-3C123_007_008.fits
-rw-r--r-- 1 azanich users   2880 Apr  4 11:50 Sum-20131210-202656-maintenance-3C123.fits
```

According to this scheme each FITS file is equivalent to one ARRAYDATA file in an MBFITS structure, see next Section.

In single-dish FITS files some information that is relevant for archive queries is stored in the FITS data tables instead of the file header (like the frequency-related information for various instrument sections). This would mean the archive has to parse through the single data files to retrieve some of the needed information. To avoid this case, the Summary file has been introduced and will contain in its header all the observation-related information which is relevant for archive queries. (The Summary file Image extension contains no data). The archive will thus have to parse only the Summary file header to retrieve all the information needed for archival searches. We anticipate here that the Summary file will not be created in the case data are written in MBFITS format where all the relevant information is stored in the various FITS headers, see end of Section 4.2.

A naming convention to univocally associate the Summary file with the correspondent scan has been established. The file name will make use of the DATE and UT fields from the scan folder name: Sum_DATEScan-UTScan-PROJID-OBJECT.fits. For instance, in the example above the Summary file name is: Sum-20131210-202656-maintenance-3C123.fits.

In the following we will generally refer to the Summary file for single-dish FITS data as 'Summary.fits'.In Appendix B a description of the Summary.fits header is given.

In some cases, typically for spectroscopic observations, the instrumental setup may change during the scan execution (f.i. frequency shift due to earth rotation of a spectral line during the observation, need to shift the band to compensate and keep the line at the band centre). It has been agreed that in any case **the Summary.fits file will always contain only the setup of the first subscan.**

4.2 HIERARCHICAL MBFITS

Initially MBFITS was developed to define a new, FITS-based, data format for multifeed receivers to be used at the IRAM 30m and APEX telescopes, in order to share software developments for calibration and data reduction. Later on, changes in the format structure and the addition of keywords needed to accommodate single-dish configurations (particularly multiple beam observing and multiple frontend/backend combinations) have been done and the MBFITS format can now be considered an independent format. It can be used for all single-dish bolometer and heterodyne observations including multiple frontend/backend combinations and array receivers. Since July 2007 the MBFITS format description is an officially registered FITS convention.

The standard FITS requires that binary tables are written sequentially in the data file. When writing an MBFITS during an observation, the underlying library (f.i. CFITSIO) continuously needs to rearrange the tables to make room for new incoming data. The alternatives of keeping the data in memory or in temporary files before writing is impracticable because of the final size of the files for long (>100 seconds) subscans which occur for wobbler or frequency switched scans or for long bolometer maps. Copying at the end would introduce large overheads and delay the online calibration and display. To remedy this problem, the FITS hierarchical grouping standard (see Jennings et al. 1997) has been employed. In this case, each MBFITS binary table is stored in a separate file inside a directory tree. This removes the problems of re-arranging the data since each single file can now simply be written sequentially.

The MBFITS hierarchical grouping directory is defined as follows:

➔ Main directory name according to the OBSDATE and PROJID keywords. Inside this main directory there are the files for the scan-level tables:

- the grouping table file: GROUPING.fits
- the scan table file: SCAN.fits
- the FEBEPAR table files for each FEBE combination: <FEBE name>-FEBEPAR.fits

➔ The actual data is stored in sub-directories per subscan named according to the subscan number. Each sub-directory contains the following types of member files:

- the MONITOR table file: MONITOR.fits
- the ARRAYDATA table files for each FEBE combination and baseband:
  <FEBE name>-ARRAYDATA-<Baseband number>.fits
- the DATAPAR table files for each FEBE combination: <FEBE name>-DATAPAR.fits

The GROUPING.fits file stores the locations of the member files so that the data set can be accessed as one entity.

The naming convention for one MBFITS directory is identical to that for FITS scan folders: DATE-UT-PROJID-OBJECT. As an example, here is the list of files present in the hierarchical MBFITS file named 20120920-135324-143917-libardi_3C286:

```
drwxr-xr-x   4096 Sep 20  2012 0001
drwxr-xr-x   4096 Sep 20  2012 0002
drwxr-xr-x   4096 Sep 20  2012 0003
drwxr-xr-x   4096 Sep 20  2012 0004
drwxr-xr-x   4096 Sep 20  2012 0005
drwxr-xr-x   4096 Sep 20  2012 0006
drwxr-xr-x   4096 Sep 20  2012 0007
```

```
drwxr-xr-x    4096 Sep 20  2012 0008
drwxr-xr-x    4096 Sep 20  2012 0009
drwxr-xr-x    4096 Sep 20  2012 0010
drwxr-xr-x    4096  Feb  1  2013 0011
drwxr-xr-x    4096  Feb  1  2013 0012
-rw-r--r--    17280 Sep 20  2012 CCC-Backend-FEBEPAR.fits
-rw-r--r--    23040 Sep 20  2012 GROUPING.fits
-rw-r--r--    17280 Sep 20  2012 SCAN.fits
```

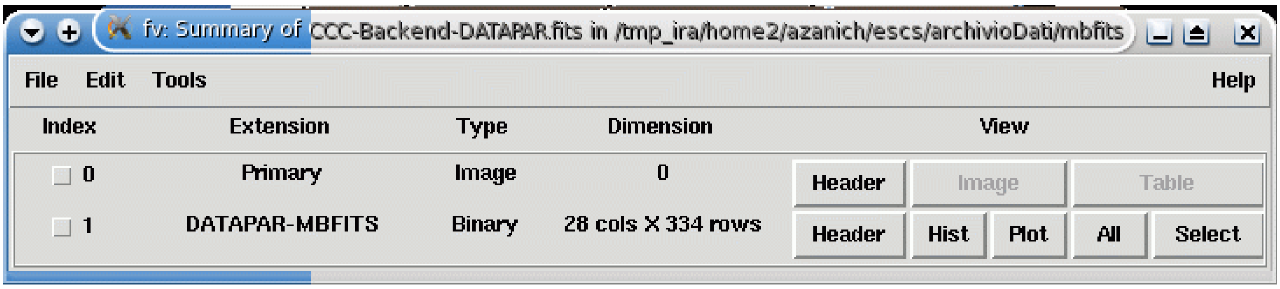
And, for each 000* sub-directory:

```
-rw-r--r--   17280 Sep 20  2012 CCC-Backend-ARRAYDATA-01.fits
-rw-r--r--   17280 Sep 20  2012 CCC-Backend-ARRAYDATA-02.fits
-rw-r--r--   86400 Sep 20  2012 CCC-Backend-DATAPAR.fits
-rw-r--r--   17280 Sep 20  2012 MONITOR.fits
```

In this example, the observation was done in cross-scan mode: crosses composed of two subscans in orthogonal direction were done on a given source. Each CCC-Backend-ARRAYDATA-01.fits contains one subscan, each directory name 000* contains one cross, and the total number of crosses performed during the observation is expressed by the number of 000* sub-directory (and thus varies from one observation to the other).

All the files composing an MBFITS are FITS files composed by a Primary HDU plus one other HDU, like the example in Figure 1.

Figure 1: summary of the content of the CCC-Backend-DATAPAR.fits file.



The Primary (HDU 0) header of the GROUPING.fits file contains some relevant keywords for the Archive, like f.i. OBJECT, EXPTIME and PROJID, while the Primary headers of the other files contain only the small set of keywords needed to define the file type as FITS and are thus not useful for the archive.

The second HDU of each fits file composing an MBFITS contains some relevant keywords related to the content of that file (f.i. the HDU 1 of the *DATAPAR.fits file contains the FEBE keyword), and the keywords needed to describe the data table contained in that file.

The complex structure of the MBFITS format cause a redundancy of keywords among the GROUPING.FITS file primary header and the HDU 1 headers of the other files, and some keywords are found in more than one header.

The hierarchical structure of MBFITS and the presence in the headers of all the information relevant for archive queries makes it unnecessary a Summary.fits like for SD data written in FITS format. From the point of view of the Archive the minimum number of HDU headers (i.e. of files in the MBFITS structure) to be inspected is defined by taking into account all the keywords that can be queried, see Sect. 7 and following.

5 SOME CONSIDERATIONS FOR AN ARCHIVE

➔ **GENERAL:** the VLBI-IT and SD formats are different in their FITS keywords set. Indeed, the two observing modes can be considered as two distinct instruments. **For FITS files coming from VLBI-IT observations the keyword TELESCOP will take the value 'VLBIT'.** This favors the design of an archive with two distinct "query trees" at the very high level: one for interferometric and the other for single-dish observations. This also allows single-dish queries to use the full SD keyword data set, while not forcing the VLBI-IT to add a huge number of keywords not foreseen in the format (and possibly even not useful).

➔ **GENERAL:** three types of users for the radio archive have been identified. The first one is the PI of a given observational project, who will have proprietary access to his data for a defined period of time, after which data will become public; the second one is a generic user who can query all the public data in the archive; the third one is the so-called archive administrator, typically someone of the observatory technical staff, who may want to query the archive for maintenance/statistics purposes**.**

➔ **GENERAL:** related to previous consideration, in the case of technical staff querying the archive it would be desirable to be able to retrieve either all the data files satisfying the query or, in some cases, the values that a query parameter has in any of the data files satisfying the query.

➔ **VLBI-IT:** the expected number of observations to be archived for the VLBI-IT antenna network (possibly also in collaboration with other antennas) is of the order of 10/year in the coming years.

➔ **SD:** the two formats for single-dish observations (FITS and hierarchical MBFITS) have slightly different keyword sets, with the former having the smaller one. An effort is being made in order to use the same name for the majority of the keywords relevant for archival search. Nevertheless, in some cases the names of a keyword hosting the same information is different in the two formats (f.i. the source Right Ascension). This does not constitute a problem for the archive, once the database tables are properly defined (see Sect. 6).

➔ **SD:** it is possible that a keyword contains a quantity written with different measurement units in FITS and MBFITS files. This happens for RA and Dec of the target: our single-dish FITS format stores them in radians while MBFITS stores them in degrees. To conform to the FITS standard convention for commonly used keywords (see http://fits.gsfc.nasa.gov/), we are planning to change RA and Dec units to degrees as well in a future version of our FITS writing code. These are the only two query parameters, among those listed in Sect. 6, for which we have detected a difference in measurement units between FITS and MBFITS.

6 ARCHIVING THE DATA

In this Section we try to summarize how data are prepared and passed to the Archive in the different cases of VLBI-IT and SD observations.

6.1 VLBI-IT DATA

Each VLBI-IT observational set to be sent to the archive will contain the following files:

- FITS Visibility File (1-50 Gby)
- Vex file (~100 KBy)
- Summary file, xml type (~100 KBy)
- Antenna log files, txt type, one per antenna involved in the observation (1 MBy)
- Correlation process Comment File, txt type (1 Kby)
- *.antab* files, one per antenna involved in the observation, containing antenna calibration information.

Note that the antenna log files contain a report of antenna operations during the observation, while the *.antab* files list information related to antenna calibration (receiver gain and system temperature, typically). In the future we plan to have an automatic pipeline for data processing that will be run on each observational data set. At that time, the archive will possibly include also data ready for the analysis (and the set of query keywords will be modified accordingly).

6.2 SD DATA

The observing system at the radio telescopes executes the observations, and output the data in FITS or MBFITS format by means of an observing schedule. The schedule is a fundamental part of each observation since it contains information on the characteristics of the observation itself and on the telescope setup. The observing system creates also a log file containing information on the system status during the observation.

Generally observers or archive users will not be interested in the type of information contained in the schedule and log files. However, there are cases in which such information may prove to be fundamental to understand the data quality. For instance, offsets in the position of a source may be caused by problems in the antenna tracking, which are recorded in the log file. For this reason, it has been decided that the archive will contain also schedules and logfiles, that can be retrieved on request.

Each schedule may command the execution of more than one observing scan, i.e. may produce more than one scan folder. To avoid duplication of data in the archive, it has been decided that the schedule and the logfile will not be attached to each scan folder but will be sent to the archive as a separate set every time the schedule itself has been fully executed or has been stopped.

The schedule itself consists of four files (scheduleName.bck, scheduleName.cfg, scheduleName.lis, scheduleName.scd) which are organized in a *schedule folder* and are used by the antenna observing system to get the information needed for the observation. To avoid duplications, the name of the schedule folder and the schedule files will mandatory contain a time tag, determined by the time at which the schedule files are copied to the observing computer to be executed. This means that the schedule time tag may be slightly different from the time tag of the sngle scans. An example of schedule folder and its content is:

```
home2/escs/archivioDati/esempio_struttura_fits/schedula> ls -la GZT06C_20140308-105028/
total 24
drwxr-xr-x 2 azanich users 4096 Mar 19 23:14 .
drwxr-xr-x 3 azanich users 4096 Mar 19 23:13 ..
-rw-r--r-- 1 azanich users  130 Jun  5  2014 GZT06C_20140308-105028.bck
-rw-r--r-- 1 azanich users  128 Mar  8  2014 GZT06C_20140308-105028.cfg
-rw-r--r-- 1 azanich users  292 Mar  8  2014 GZT06C_20140308-105028.lis
-rw-r--r-- 1 azanich users  978 Mar  8  2014 GZT06C_20140308-105028.scd
```

The schedule name time tag will also be different from the logfile name time tag: in the previous example the logfile name would be for instance: GZT06C_20140308-105930.log .

The schedule and log files are associated to the data by means of keywords in the Summary.fits header (for FITS format) or in the GROUPING.FITS header (for MBFITS format).

The flow from the telescope to the archive will be something like this:

a) the observing system performs the observations and creates data in FITS or MBFITS format as well as a logfile on the basis of a given schedule.
b) A dedicated script, a so called *Finalizer*, will take care of packing data into tar files, one for each observing scan. The script will also create the tar file containing the schedule folder and the logfile once the execution of the schedule is completed.
c) The archive will receive in input the tar files mentioned above.

The archive will thus receive data organized in tar files, each one containing

- one FITS or MBFITS scan folder, or
- schedule folder + logfile

In this way, from the execution of an observing schedule the archive will receive in input: *n* scan folders created by the schedule execution, passed as *n* tar files, plus one tar file contaning the schedule itself and its relative logfile.

7 VLBI-IT DATA: PARAMETERS FOR THE ARCHIVE

In this Section we list the keywords to be used for archival search of VLBI-IT observation.

The Summary.xml file is the reference file for the archive in the case of VLBI-IT data sets and contains *at least* the keywords to be offered for archival search of VLBI-IT data, namely:

- project ID of the observations;
- PI name;
- source name;
- source RA and DEC;
- date and time of the observation;
- italian antennas participating in the array;
- minimum and maximum frequency (observing band);
- spectral channels and resolution;
- data rate;
- time spent on-source.

Table 2 lists the keywords relative to one source in a VLBI-IT dataset that are relevant for archive queries with details on their measurement units and a brief description. Note that, contrary to what happens for SD observations, the Summary.xml file for VLBI-IT data will contain many of these keyword sets (one for each of the observed sources in the VLBI-IT dataset) and will have to be parsed accordingly. See Appendix A for an example of a Summary.xml file.

Query parameter names are those that should appear in the archive web form, to be queried by users. For this reason they often differ from the correspondent keyword name since they do not suffer from the limitations and conventions of the FITS keywords.

Table 2: VLBI-IT keywords to be used for archive queries.

Query parameter Name	**Summary.xml keyword name**	**Keyword Units**	**Description**
PROJECT ID	PROJID		Project ID
TELESCOPE	TELESCOP		Telescope name
PI NAME	PINAME		Name of the PI of the project
TARGET	OBJECT(note A)		Target source name
RA	RA (note A)	deg	Target Right Ascension
DEC	DEC (note A)	deg	Target Declination
OBSERVATION DATE	START		Start date of observation
ANTENNAS	ANTENNAS (note B)		Italian antennas participating in the array
MINIMUM FREQUENCY	FREQ_MIN (note C)	MHz	Minimum observing frequency
MAXIMUM FREQUENCY	FREQ_MAX (note C)	MHz	Maximum observing frequency
SPECTRAL CHANNELS	CHANNELS		Number of spectral channels
SPECTRAL RESOLUTION	CH_RESOL	MHz	Spectral resolution
DATA RATE	RATE	Mbit/sec	Data rate
TIME ON SOURCE	ONTIME	sec	Time on-source

Note A:

When a user wants to make a query on a specific source using its name or coordinates, the archive will return all the datasets containing that source, without splitting the dataset content (see Sect. 3). This means that the user retrieves from the archive one (or more) full dataset(s), and he may then split the data at home according to his needs.

Note B:

The VLBI-IT network may include the Italian antennas as well as radio telescopes from other counties. Each telescope has its unique naming code, being 'Mc', 'Nt' and 'Sr' for Medicina, Noto and SRT respectively. In general, the antennas participating in the VLBI-IT interferometer may vary in each observation. It has been agreed that the archive may be queried on the presence of at least one of the three Italian radio telescopes in an observation, by means of the query parameter ANTENNAS.

Note C:

When a user wants to make a query on frequency, he selects the minimum and maximum frequency he is interested in and the query will return all those datasets for which the [FREQ_MIN, FREQ_MAX] range overlaps with the range [MINIMUM FREQUENCY, MAXIMUM FREQUENCY].

As said in Sect. 6.1, it is foreseen that at some point the archive will include also pipeline processed data. In that case, the set of query keywords will be modified accordingly.

Table 3 summarizes the formats, units and values for the parameters that are queried in the archive web form (from Table 2) and the allowed values of the associated VLBI-IT keyword. Columns 1 to 3 list the query parameter name as well as the format and units in which it should be showed in the web form. Column 4 lists the possible values that each parameter can have: some of them are user-selected (i.e. no pre-defined set of values is offered) while some others may assume a limited range of values, mainly set by the keyword values themselves. Column 5 and 6 list the values that the associated VLBI-IT FITS keywords may have and their format.

The query on RA, Dec should include a search radius around the position (if radius=0.0, exact match). Searches in OBSERVATION DATE should require a minimum and maximum value of the query parameter.

The query parameter values for spectral channels, spectral resolution and data rate are to be intended as minimum values, that is the returned dataset are those having the selected number of channels or higher, and/or the selected resolution or better (smaller values), and/or the selected data rate or higher.

Table 3: Allowed values for query parameters in the archive web form for VLBI-IT data.

Query parameter	Query parameter format	Query parameter units	Allowed values for query parameter	Keyword values in VLBI-IT FITS files	Keyword Format/Unit
PROJECT ID	String	n/a	User-selected	String	String
TELESCOPE	String	n/a	'VLBI-IT'	'VLBIT'	String
PI NAME	String	n/a	User-selected	Any	String
TARGET	String	n/a	User-selectd	Any	String
RA	hh:mm:ss.sss or decimal degs	hh:mm:ss.sss or decimal degs	User-selected with search radius	Any allowed value in degrees	Double
DEC	dd:mm:ss.ss or decimal degs	dd:mm:ss.ss or decimal degs	User-selected with search radius	Any allowed value in degrees	Double
OBSERVATION DATE	String	n/a	Search in a range between DATE1 and DATE2	Any value in the form: 'YYYY-MM-DD hh:mm:ss'	String
ANTENNAS	String	n/a	User-selected	'Mc', 'Nt', 'Sr'	String
MINIMUM FREQUENCY	Float	MHz	User-selected	Any value in MHz	Float
MAXIMUM FREQUENCY	Float	MHz	User-selected	Any value in MHz	Float
SPECTRAL CHANNELS	Integer	n/a	User-selected	Any	Integer
SPECTRAL RESOLUTION	Float	MHz	User-selected	Any value in MHz	Float
DATA RATE	Float	Mbit/sec	User-selected	Any	Float
TIME ON SOURCE	Float	sec	User-selected	Any	Integer

8 SD DATA: PARAMETERS FOR THE ARCHIVE

In this Section we describe the set of keywords, common to FITS and MBFITS files, that have been identified as interesting for archive queries.

Table 4 lists the FITS files keywords relevant for archive queries with details on their measurement units and a brief description. All these keywords are in the Primary Header of the Summary.fits file associated to each scan.

It has been agreed that a dedicated keyword containing the name of the project PI (PINAME) is included in both FITS and MBFITS single-dish data. This keyword value will be filled as soon as an automated procedure to communicate the PI name from the Time Allocation Committee/ Scheduler to the observing system will be available. Until then, this parameter will not be available for archival queries on SD data and for this reason it is not included in the following tables, while it is listed among the keywords of the Summary.fits file in Appendix C.

Table 4: FITS keywords to be used for archive queries on SD data.

Query parameter name	**FITS keyword name**	**FITS Units**	**Description**
TELESCOPE	TELESCOP		Telescope name
FRONTEND	HIERARCH ReceiverCode		Receiver name
BACKEND	HIERARCH BackendName		Backend name
MINIMUM FREQUENCY	FREQ*n* (note A)	MHz	Minimum of the desired frequency range
MAXIMUM FREQUENCY	FREQ*n* (note A)	MHz	Maximum of the desired frequency range
BANDWIDTH	BWID*n* (note A)	MHz	Band width
TARGET	OBJECT		Target source name
OBSERVATION DATE	DATE-OBS (note A)		Summary File creation date
RA	HIERARCH RightAscension	rad	Target Right Ascension
DEC	HIERARCH Declination	rad	Target Declination
EQUINOX	EQUINOX		Equinox of RA, Dec
EXPOSURE TIME	EXPTIME	sec	Total integration time
PROJECT ID	PROJID		Project ID
LST	LST (Note B)		Local Sidereal Time
SCAN TYPE	SCANTYPE		Scan astronomical type
SCAN MODE	SCANMODE		Mapping mode
SCAN GEOMETRY	SCANGEOM		Scan geometry
SWITCH MODE	SWTCHMOD		Switch mode
FREQUENCY RESOLUTION	FREQRES*n* **(Note C)**	Hz	Frequency resolution
REST FREQUENCY	RESTFREQ*n* **(Note C)**	Hz	Rest frequency of the observed spectral line
SCAN FRAME	USRFRAME		Description of user frame
FREQUENCY BIN	CHANn		Number of spectral channels for baseband
SCAN SPEED	SCANXVEL	deg/min	Tracking rate (optional, OTF)
WOBBLER	WOBUSED		Wobbler used?
OBSERVER ID	OBSID		Observer or operator initials
CREATOR	CREATOR		Software (incl. version)
FORMAT VERSION	FITSVER		FITS version

Note A:
The format of the FITS keyword DATE-OBS is YYYY-MM-DDThh:mm:ss.sss, common to MBFITS definitions (see ref (1) page 31).

Note B:
The format of the single-dish FITS keyword LST is hh:mm:ss.s. This is different than for MBFITS, for which the standard LST unit is seconds.

Note C:
Frequency, bandwidth, number of spectral channels in the band, spectral resolution of each channel and rest frequency of a line for a given scan are given in the Summary.fits header as sets of values, see an example in Appendix B. For each instrumental section, groups of values (FREQ*n*, BWID*n,* CHAN*n,* FREQRES*n,* RESTFREQ*n*) are given where FREQ*n* is the starting frequency of the bandwidth BWID*n* for section *n*. The total number *n* of sections (and thus of groups of values) used in a scan is given by the Summary.fits keyword NUSEBAND. The values RESTFREQ*n* are foreseen in case the observer wants to center each spectral section on a particular spectral line. The query parameter values for spectral channels and spectral resolution are to be intended as minimum values, that is the returned dataset are those having the selected number of channels or higher, and/or the selected resolution or better (smaller values).

When a user wants to make a query on frequency, he selects the minimum and maximum frequency he is interested in and the query will return all those scans for which *at least* one of the FREQn values fall inside the range [MINIMUM FREQUENCY, MAXIMUM FREQUENCY]. This is done as follows:

1) the user selects MINIMUM FREQUENCY, MAXIMUM FREQUENCY which define the Query_Range.
2) The database reads for each scan the couples of starting frequencies and bandwidths (FREQ*n*, BWID*n*) and computes for each section the values OBS_MIN_FREQUENCY*n*=FREQ*n* and OBS_MAX_FREQUENCY*n*=FREQ*n*+BWID*n* which define the Observed_Range.
3) The archive returns all those scans for which there is a superposition between the Query_Range and the Observed_Range.

To make a practical example, let's suppose the user selects MINIMUM FREQUENCY=22000 MHz and MAXIMUM FREQUENCY=24000 MHz, so that Query_Range=[22000,24000]. Let's suppose also that in the archive there is a scan which contains a section (n=1 in this example) with FREQ1=21900 MHz and BWID1=1000 MHz, thus Observed_Range=[21900,22900]. In this example, there is a superposition between Query_Range and Observed_Range and this scan must be returned to the user as a query match.

When a user wants to make a query on bandwidth, an *exact match* between the query parameter BANDWIDTH and the database values for BWID*n* is to be done.

Table 5 lists the same keyword set as Table 4 but for MBFITS files. Note that Table 5 is to be taken as preliminary since the MBFITS format is still under test (especially for the spectroscopic part) and some things may still change. Column 4 lists the header where the specific keyword is to be searched in the hierarchical MBFITS structure.

Query parameter names are those that should appear in the archive web form, to be queried by users.

With regards to data structure, it is foreseen to include the keywords related to frequency also in the ARRAYDATA files of the MBFITS structure. This affects how/where the archive looks for the relevant parameters but does not impact on the query form itself.

Table 5: MBFITS keywords to be used for archive queries on SD data (preliminary description).

Query parameter name	**MBFITS keyword name**	**MBFITS units**	**Where in MBFITS**	**Description**
TELESCOPE	TELESCOP		GROUPING HDU 0	Telescope name
FRONTEND	**(Note A)**			Receiver name
BACKEND	**(Note A)**			Backend name
MINIMUM FREQUENCY	**(Note A)**			Minimum of the desired frequency range

Query parameter name	MBFITS keyword name	MBFITS units	Where in MBFITS	Description
MAXIMUM FREQUENCY	**(Note A)**			Maximum of the desired frequency range
BANDWIDTH	**(Note A)**			Bandwidth
TARGET	OBJECT		GROUPING HDU 0	Target source name
OBSERVATION DATE	DATE-OBS		GROUPING HDU 0	Date of observation
RA	RA	deg	GROUPING HDU 0	Target Right Ascension
DEC	DEC	deg	GROUPING HDU 0	Target Declination
EQUINOX	EQUINOX		GROUPING HDU 0	Equinox of RA, Dec
EXPOSURE TIME	EXPTIME	sec	GROUPING HDU 0	Total integration time
PROJECT ID	PROJID		GROUPING HDU 0	Project ID
LST	LST (Note B)		SCAN HDU 1	Local apparent Sidereal Time
SCAN TYPE	SCANTYPE		SCAN HDU 1	Scan astronomical type
SCAN MODE	SCANMODE		SCAN HDU 1	Mapping mode
SCAN GEOMETRY	SCANGEOM		SCAN HDU 1	Scan geometry
SWITCH MODE	SWTCHMOD		FEBEPAR HDU 1	Switch mode
FREQUENCY RESOUTION	FREQRES (**Note C**)	Hz	ARRAYDATA HDU 1	Frequency resolution
REST FREQUENCY	RESTFREQ	Hz	ARRAYDATA HDU 1	Rest freq. of observed line
SCAN FRAME	USRFRAME		MONITOR HDU 1	Description of user frame
FREQUENCY BIN	**(Note C)**		ARRAYDATA HDU 1	Number of spectral channels for baseband
SCAN SPEED	SCANXVEL	deg/min	SCAN HDU 1	Tracking rate (optional, OTF)
WOBBLER	WOBUSED		SCAN HDU 1	Wobbler used?
OBSERVER ID	OBSID		SCAN HDU 1	Observer or operator initials
CREATOR	CREATOR		GROUPING HDU 0	Software (incl. version)
FORMAT VERSION	MBFITSVER		GROUPING HDU 0	FITS version

Note A: the query parameters FREQUENCY, BANDWIDTH, FRONTEND and BACKEND have not a direct correspondence with keywords in MBFITS. As an example: FRONTEND and BACKEND correspond to only one, composite keyword in MBFITS (the FEBE keywords) that must thus be parsed for a match. The meaning of the frequency keyword in MBFITS is slightly different than in FITS: in MBFITS what is given is the central frequency of the given bandwidth, thus the search for a match should be done inside the range [frequency – bandwidth/2 ; frequency + bandwidth/2]. Exact rules for the matching of these keywords will be detailed in a future version of this document.

Note B: the format of the MBFITS keyword LST is seconds. This is different than for single-dish FITS for which the standard LST unit is hh:mm:ss.s.

Note C: MBFITS does not contain a single keyword for the query parameter FREQUENCY BIN, butmatching values must be searched by parsing data tables (each instrumental baseband has its own value for this parameter). The query parameter values for spectral channels and spectral resolution are to be intended as minimum values, that is the returned dataset are those having the selected number of channels or higher, and/or the selected resolution or better (smaller values).

With respect to previous versions of this document, now the keyword for the schedule name is no more a search option. The query form should contain a check button, named “Retrieve schedules/log files” and unselected by default, that the archive user may highlight if he wants to retrieve also schedules and log files from his query. Internally, the archive will make use of the keywords HIERARCH ScheduleName and HIERARCH Logfile Name to match data with the appropriate files (see Appendix B and C).

The keywords HIERARCH ScheduleName and HIERARCH Logfile Name are not foreseen in the MBFITS standard format and are added as HIERARCH keyword in Medicina/Noto/SRT MBFITS.

Table 6 summarizes the formats, units and values for the parameters that are queried in the archive web form (from Tables 4 and 5) and the allowed values of the associated FITS keyword. Columns 1 to 3 list the query parameter name as well as the format and units in which it should be showed in the web form. Column 4 lists the possible values that each parameter can have: some of them are user selected (i.e. no pre-defined set of values is offered) while some others may assume a limited range of values, mainly set by the keyword values themselves. For instance, the query parameter TELESCOPE can assume the values 'Any', 'MED' or 'NT': in this case, the archive user can query the TELESCOPE parameter selecting inside this pre-defined list of values. Column 5 and 6 list the values that the associated FITS/MBFITS keywords may have and their format in the FITS/MBFITS headers/tables.

The exact values allowed for the various keywords may slightly change with respect to those listed in Table 6, especially for the Noto antenna, The database will be constantly kept updated with respect to changes in the parameter values.

If possible, the query on RA, Dec should include a search radius around the position (if radius=0.0, exact match), see also Appendix D. Searches in LST and OBSERVATION DATE should require a minimum and maximum value of the query parameter.

Table 6: Allowed values for query parameters in the archive web form for SD data.

Query parameter	**Query parameter format**	**Query parameter units**	**Allowed values for query parameter**	**Keyword values in FITS/MBFITS files**	**Keyword Format**
TELESCOPE	String	n/a	'Any' + all values in col. 5	'MED', 'NT'	String
FRONTEND	String	n/a	'Any' + all those in col. 5	'LHP', 'LLP', 'SSP', 'SXP', 'CCC', 'CHC', 'XXP', 'SXP', 'KKC'	String
BACKEND	String	n/a	'Any' + all values in col. 5	'TP', 'XARCOS', 'DBBC', 'MARKVb', 'MARK5c'	String
MINIMUM / MAXIMUM FREQUENCY	Float	MHz	User-selected	Any	Float
BANDWIDTH	Float	MHz	User-selected	Any	Float
TARGET	String	n/a	User-selected	Any	String
OBSERVATION DATE	String	n/a	Search in a range between DATE1 and DATE2	Any value in the form: 'YYYY-MM-DDThh:mm:ss.s'	String
RA	hh:mm:ss.sss or decimal degs	hh:mm:ss.sss or decimal degs	User-selected with search radius	Any allowed value in units of radians/degrees (FITS/MBFITS)	Double
DEC	dd:mm:ss.ss or decimal degs	dd:mm:ss.ss or decimal degs	User-selected with search radius	Any allowed value in units of radians/degrees (FITS/MBFITS)	Double
EQUINOX	Integer	n/a	Values in col. 5	2000, 1950	Integer
EXPOSURE TIME	Float	sec	User-selected	Any	Float
PROJECT ID	String	n/a	User-selected	Any	String
LST	hh:mm:ss.s	n/a	Search in a range between LST1 and	Any	Double

Query parameter	Query parameter format	Query parameter units	Allowed values for query parameter	Keyword values in FITS/MBFITS files	Keyword Format
			LST2		
SCAN TYPE	String	n/a	'Any' + all values in col. 5	'POINT', 'FOCUS', 'CAL', 'FLUXCAL', 'SKYDIP', 'ON', 'ONOFF', 'MAP', 'FOCUS-X', 'FOCUS-Y', 'FOCUS-Z', 'FOCUS-XTILT', 'FOCUS-YTILT', 'FOCUS-ZTILT', 'UNKNOWN'	String
SCAN MODE	String	n/a	'Any' + all values in col. 5	'SAMPLE', 'RASTER', 'OTF', 'DRIFT', 'NONE', 'BEAMPARK', 'UNKNOWN'	String
SCAN GEOMETRY	String	n/a	'Any' + all values in col. 5	'SINGLE', 'LINE', 'CROSS', 'RECT', 'QUAD', 'CIRC', 'ARC', 'CURVE', 'UNKNOWN'	String
SWITCH MODE	String	n/a	'Any' + all values in col. 5	'TOTP', 'FSW', 'BEAMSW', 'HORNSW', 'LOADSW', 'CAL'	String
FREQUENCY RESOLUTION	Float	Hz	User-selected	Any	Float
REST FREQUENCY	Float	Hz	User-selected	Any	Float
SCAN FRAME	String	n/a	'Any' + all values in col. 5	'EQEQEQ', 'EQHOHO', 'HOHOHO', 'GAGAGA', 'GAHOHO'	String
FREQUENCY BINS	Integer	n/a	User-selected	Any	Integer
SCAN SPEED	Float	deg/min	User-selected	Any	Float
WOBBLER	String	n/a	'Any' + all values in col. 5	'F', 'T'	String
OBSERVER	String	n/a	User-selected	Any	String
CREATOR	String	n/a	User-selected	Any	String
FORMAT VERSION	String	n/a	User-selected	Any	String

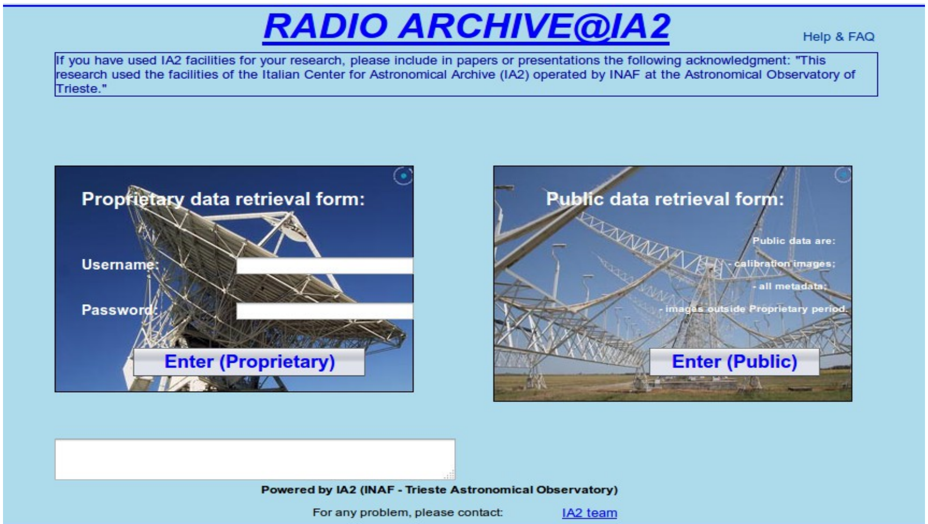
9 RADIO ARCHIVE WEB INTERFACE: SEARCH PARAMETERS

The Radio Archive users will perform dedicated queries over the Archive depending on Instrument (Telescope) features. In particular, SD and VLBI-IT observation shall share a generic subset of parameters but should be specialized for each one in terms of selection parameters and query outputs.
Common features are for example the user authentication page, the object or coordinates characteristics, the frequencies fields. Specialized fields shall be related to the Observation mode, the frontend – backend configuration and so on. Following a sort of Mock-up of the Radio web Interface is presented.

9.1 ARCHIVE WEB ACCESS PAGE

Medicina and Noto Radio Archive will be accesible in two ways: privately or publicly, see Figure 2. Public data follows the INAF policy and are available for download without any requirement of registration to everyone. Private data are available to Principal Investigator or authorized person already registered into the local Authentication and Authorization environment. A Single Sign On mechanism is foreseen but not immediately implemented.
Authentication and Authorization will be based on PI (or delegated) username and password directly connected to the National Astrophysic Institute registry. People not afferring to INAF will be added to a local identity provider if the case. Non-authorized private access will be simply rejected and an error message written into the text field zone.
In both public or private cases, a query form will be presented to the user. A common subset of parameters will be available to perform the query with the possibility to choose the instrument (Medicina, Noto) or type of observation (SD or VLBI).
Specific query could be performed opening -with a button- a hidden section, depending on observation type (SD or VLBI). Once the selection parameters are set and button “SEARCH” is clicked, the result set is presented to the user. Additional informations check box are available like Add Schedule, Add Logs to the downloadable files in the result set page.
Result set populate dedicated tables. Result set columns are composed by checking the checkbox corrispondent to interesting parameters and filters are applied once parameters value are inserted into the text fields. A list of data descriptors (metadata) that satisfy the query are returned.

Figure 2: Archive Web access page



General INAF policy states all metadata are public while only owners could access private files. One of the result set table colum is the link to file and it is available only if the authorization condition is satisfied. Clicking on the link to file, the user can download the file immediately or checking the corrispondent checkbox can select the file for succeeding download. Single file format is a tar folder while multiple file selection redirect to a next page in which the result of file aggregation into a tar file is presented.
In Appendix E diagrams describing the database architecture, the software configuration model and the data model are presented.

APPENDIX A. VLBI-IT SUMMARY.XML FILE

In the following an example of the Summary.xml file content for a VLBI-IT observation is described.


```
<?xml version='1.0' encoding='ASCII'?>
<database>
  <table>
    <PROJID>1313a</PROJID>
    <TELESCOP>VLBIT</TELESCOP>
    <PINAME>Hayley Bignall</PINAME>
    <OBJECT>0212+735</OBJECT>
    <RA>30.2918926036</RA>
    <DEC>73.8257282611</DEC>
    <START>2014-03-22 14:43:19</START>
    <ANTENNAS>Nt-Md-Mc</ANTENNAS>
    <FREQ_MIN>4966.49</FREQ_MIN>
    <FREQ_MAX>5094.49</FREQ_MAX>
    <CHANNELS>16.0</CHANNELS>
    <CH_RESOL>8.0</CH_RESOL>
    <RATE>512.0</RATE>
    <ONTIME>4426</ONTIME>
  </table>
  <table>
    <PROJID>1313a</PROJID>
    <TELESCOP>VLBIT</TELESCOP>
    <PINAME>Hayley Bignall</PINAME>
    <OBJECT>0340+362</OBJECT>
    <RA>45.7247090019</RA>
    <DEC>36.370119325</DEC>
    <START>2014-03-22 14:43:19</START>
    <ANTENNAS>Nt-Md-Mc</ANTENNAS>
    <FREQ_MIN>4966.49</FREQ_MIN>
    <FREQ_MAX>5094.49</FREQ_MAX>
    <CHANNELS>16.0</CHANNELS>
    <CH_RESOL>8.0</CH_RESOL>
    <RATE>512.0</RATE>
    <ONTIME>4427</ONTIME>
  </table>
  <table>
```

<PROJID>1313a</PROJID>
<TELESCOP>VLBIT</TELESCOP>
<PINAME>Hayley Bignall</PINAME>
<OBJECT>0507+179</OBJECT>
<RA>75.1673247586</RA>
<RA>75.1673247586</RA>
<DEC>18.01155045</DEC>
<START>2014-03-22 14:43:19</START>
<ANTENNAS>Nt-Md-Mc</ANTENNAS>
<FREQ_MIN>4966.49</FREQ_MIN>
<FREQ_MAX>5094.49</FREQ_MAX>
<CHANNELS>16.0</CHANNELS>
<CH_RESOL>8.0</CH_RESOL>
<RATE>512.0</RATE>
<ONTIME>4428</ONTIME>
</table>
<table>
<PROJID>1313a</PROJID>
<TELESCOP>VLBIT</TELESCOP>
<PINAME>Hayley Bignall</PINAME>
<OBJECT>0642+449</OBJECT>
<RA>90.7755627775</RA>
<DEC>44.8546083583</DEC>
<START>2014-03-22 14:43:19</START>
<ANTENNAS>Nt-Md-Mc</ANTENNAS>
<FREQ_MIN>4966.49</FREQ_MIN>
<FREQ_MAX>5094.49</FREQ_MAX>
<CHANNELS>16.0</CHANNELS>
<CH_RESOL>8.0</CH_RESOL>
<RATE>512.0</RATE>
<ONTIME>4426</ONTIME>
</table>
</database>

APPENDIX B. SINGLE-DISH FITS HEADER

Table B.1 shows an example of FITS Primary HDU from a Single-Dish observations. With respect to previous versions some keywords have been moved to the Summary.fits header (see Appendix C) and some keyword names are changed to homogeneize as far as possible the FITS and MBFITS header structures.

Table A.1: Example Primary HDU content. Numerical values have to be taken as examples.

SIMPLE	T	file does conform to FITS standard
BITPIX	8	number of bits per data pixel
NAXIS	0	number of data axes
EXTEND	T	FITS data set may contain extensions
DATE	'2013-11-15T16:38:13.9'	file creation date (YYYY-MM-DDThh:mm:ss UT)
HIERARCH Project_Name	'SRT-SCICOM_Test_3C295_K_2'	Name of the project
OBSID	' '	Name of the observer
TELESCOPE	'SRT '	Name of the station
HIERARCH SiteLongitude	0.161358481873679	Longitude of the site (radians)
HIERARCH SiteLatitude	0.689283579621821	Latitude of the site (radians)
HIERARCH SiteHeight	650.	Height of the site (meters)
BEAMS	7	Number of beams
SECTIONS	14	Total number of sections
HIERARCH Sample Size	4	Number of bytes of a data
HIERARCH ReceiverCode	'KKG '	Keyword that identifies the receiver
OBJECT	'3C295 '	Source identifier
HIERARCH RightAscension	3.71468605789256	Source right ascension at J2000 (radians)
HIERARCH Declination	0.91110550277234	Source declination at J2000 (radians)
VLSR (note A)	0.	Source radial velocity
HIERARCH Azimuth Offset	0.	Longitude offset in horizontal frame
HIERARCH Elevation Offset	0.	Latitude offset in horizontal frame
HIERARCH RightAscension Offset	0.	Longitude offset in equatorial frame
HIERARCH Declination Offset	0.00148352986419518	Latitude offset in equatorial frame
HIERARCH GalacticLon Offset	0.	Longitude offset in galactic frame
HIERARCH GalacticLat Offset	0.	Latitude offset in galactic frame
SCANID	1	Scan Identifier
HIERARCH SubScanID	2	Subscan Identifier
HIERARCH ScheduleName	'/archive/schedules/scicom/SRT-SCICOM_Test_3C295_K_2.sc'	Schedule Name
HIERARCH LogfileName	'/archive/logfiles/scicom/SRT-SCICOM_Test_3C295_K_2.log'	Logfile Name
HIERARCH SubScanType	'RA '	describes the scan type
HIERARCH Scan Tag	1	Scan tag identifier
TSYS	39.1	System temperature
TSYSTIME (note B)	2013-115-16-38-18.0'	Measurement date-time of Tsys (YEAR-DOY-HH-MM-SS.Ss)

Note A: The VLSR spectroscopic keyword will be substituted by two keywords, one for the velocity value and the other for the velocity system. In principle it is possible to have more than one couple of these keywords, according to the used backend sections (see also note (A) in Appendix C). The coding/naming convention of these spectroscopic keywords is ongoing.

Note B: TSYSTIME format has a syntax different from the "classical" DATE format. It is in fact based on the day-of-year: YEAR-DOY-HH-MM-SS.s. Could it be that we align it to other date formats? (that is YYYY-MM-DDThh:mm:ss.s)

APPENDIX C. SINGLE-DISH *SUMMARY.FITS* FILE HEADER

Here we list the content of the primary header of the Summary.fits file described in Section 4.1 for Single-Dish observations.

The keywords NUSEBAND, FREQ*n*, BWID*n* are to be used together for archival queries on the frequency, see Section 7. The number *n* can take integer values from 1 to NUSEBAND, where NUSEBAND describes the number of sections used in that observation. Each section can have in principle different values for frequency, bandwidth and number of spectral channels.

Note that the keyword PROJID in Summary.fits is not equivalent to HIERARCH Project_Name in the FITS main header. PROJID is the project unique code assigned by the TAC and used by the archive to couple a Principal Investigator to a given dataset and its proprietary period. The keyword HIERARCH Project_Name is related to a particular schedule and contains a more detailed description of the project. For instance, inside the same project (PROJID = 001-23D) both survey type (HIERARCH Project_Name="KNoWS survey") and follow-up type (HIERARCH Project_Name="KNoWS followup") observatons may be required, using different schedules.

Table C.1: Example Primary HDU for a Summary.fits file in a SD observation**. In green, keywords that changed name (or part of it) since the last version of this document.**

FITS keyword name	**Value**	**Description**
TELESCOP	'MED'	Telescope name
PINAME	'Mack'	Name of the Principal Investigator
HIERARCH ReceiverCode	'CCC'	Receiver name
HIERARCH BackendName	'TP'	Backend name
OBJECT	'3C123'	Target source name
DATE-OBS (note A)	'2013-09-20T13-53.25.084'	Summary File creation date
HIERARCH RightAscension	1.20895868780092	Target Right Ascension
HIERARCH Declination	0.517848007827577	Target Declination
EQUINOX	2000	Equinox of RA, Dec
EXPTIME	60.0	Total integration time
PROJID	'13-10'	Project ID
LST	'12:23:58.9'	Local Sidereal Time
SCANTYPE	'ONOFF'	Scan astronomical type
SCANMODE	'NONE'	Mapping mode
SCANGEOM	'SINGLE'	Scan geometry
SWTCHMOD	'TOTP'	Switch mode
USRFRAME	'HOHOHO'	Description of user frame
HIERARCH ScheduleName	'k3c123.scd'	Schedule name
HIERARCH LogfileName	SRT-SCICOM_Test_3C295_K_2.log'	Logfile name
SCANXVEL	0.0	Tracking rate (optional, OTF)
WOBUSED	'F'	Wobbler used?
OBSID	'Righini'	Observer or operator initials
CREATOR	'ESCS v0.3'	Software (incl. version)
FITSVER	'1.0'	FITS version

FITS keyword name	**Value**	**Description**
NUSEBAND	3	Number of sections
FREQ1	4800.0	Frequency of the n-th section
BWID1	150.0	Bandwidth of the n-th section
CHAN1	0	N. of spectral channels for the n-th section.
FREQRES1	0.0	Frequency resolution
RESTFREQ1	0.0	Rest freq. of observed line
FREQ2	5000.0	Frequency of the n-th section
BWID2	330.0	Bandwidth of the n-th section
CHAN2	0	N. of spectral channels for the n-th section.
FREQRES2	0.0	Frequency resolution
RESTFREQ2	0.0	Rest freq. of observed line
FREQ3	5200.0	Frequency of the n-th section
BWID3	330.0	Bandwidth of the n-th section
CHAN3	0	N. of spectral channels for the n-th section.
FREQRES3	0.0	Frequency resolution
RESTFREQ3	0.0	Rest freq. of observed line
SCANOFF	0.0	Position Offset (deg)
SCANSPAN **(Note C)**	0.0	Scan span (deg)
HIERARCH SCANSTART **(Note D)**	0.00000000	Scan starting position (deg)

Note A:
The format of the FITS keyword DATE-OBS is YYYY-MM-DDThh:mm:ss.sss, common to MBFITS definitions (see ref (1) page 31).

Note B:
The format of the single-dish FITS keyword LST is hh:mm:ss.s. This is different than for MBFITS, for which the standard LST unit is seconds.

Note C: a possible change of this keyword name from SCANSPAN to SCANLEN is foreseen to homogeneize with MBFITS keyword SCANLEN, defined as the user-defined line length for OTF and Raster observing modes.

Note D: This keyword value may be changed to INITPOS.

APPENDIX D. SEARCHING BY RADIUS AROUND A GIVEN POSITION

The Archive user may look for a source which is not present as target of an observation, but which has however been observed in a map done around another target.

Here we describe how the size of a map can be computed starting from the keywords present in an MBFITS, so that the Archive can compute map sizes and check if a given coordinate falls inside one of the available maps

The following holds true for MBFITS data.
Keywords containing information on the scan geometry are found in the SCAN.fits file.
There are two kinds of maps that can be executed: On-The-Fly (continuous acquisition of data while moving along a scan) or Raster (acquisition at specified positions while moving along a scan).
The size of a map is computed in different ways depending on the observing mode:

- On-The-Fly map:

map size = SCANLEN, where the SCANLEN keyword is found in HDU 1 of SCAN.fits

- Raster map:

map size is determined by means of two keywords in HDU 1 of SCAN.fits: SCANLEN (that in Raster case represent the n. of points to be acquired in a subscan), SCANXSPC, SCANYSPC (step between scan/raster lines). We have:
map size X = SCANLEN * SCANXSPC
map size Y = SCANLEN * SCANYSPC

To determine the absolute coordinates of map corners, one needs to know where the map is starting with respect to the source coordinates (top left corner, upper left corner...) i.e. to determine offsets.

The computed coordinates will refer to the coordinate frame in which the map has been executed. Maps (and subscans in general) can be performed along Az/El direction, RA/Dec, Galactic Lon/Lat. If the archive is going to adopt RA,Dec as the reference for coordinate search, when searching for a position in an archived map the proper conversion (if needed) between search coordinates and map execution coordinates must be performed. The keyword for the map frame is SCANDIR, again in SCAN.fits HDU1 header.

APPENDIX E. ARCHIVER AND DATABASE ARCHITECTURE

Figure E1: foreseen architecture used to ingest data, independently from data format (FITS or MBFITS). All the metadata informations are stored in the same or in different tables, depending on software configuration and on telescope site.

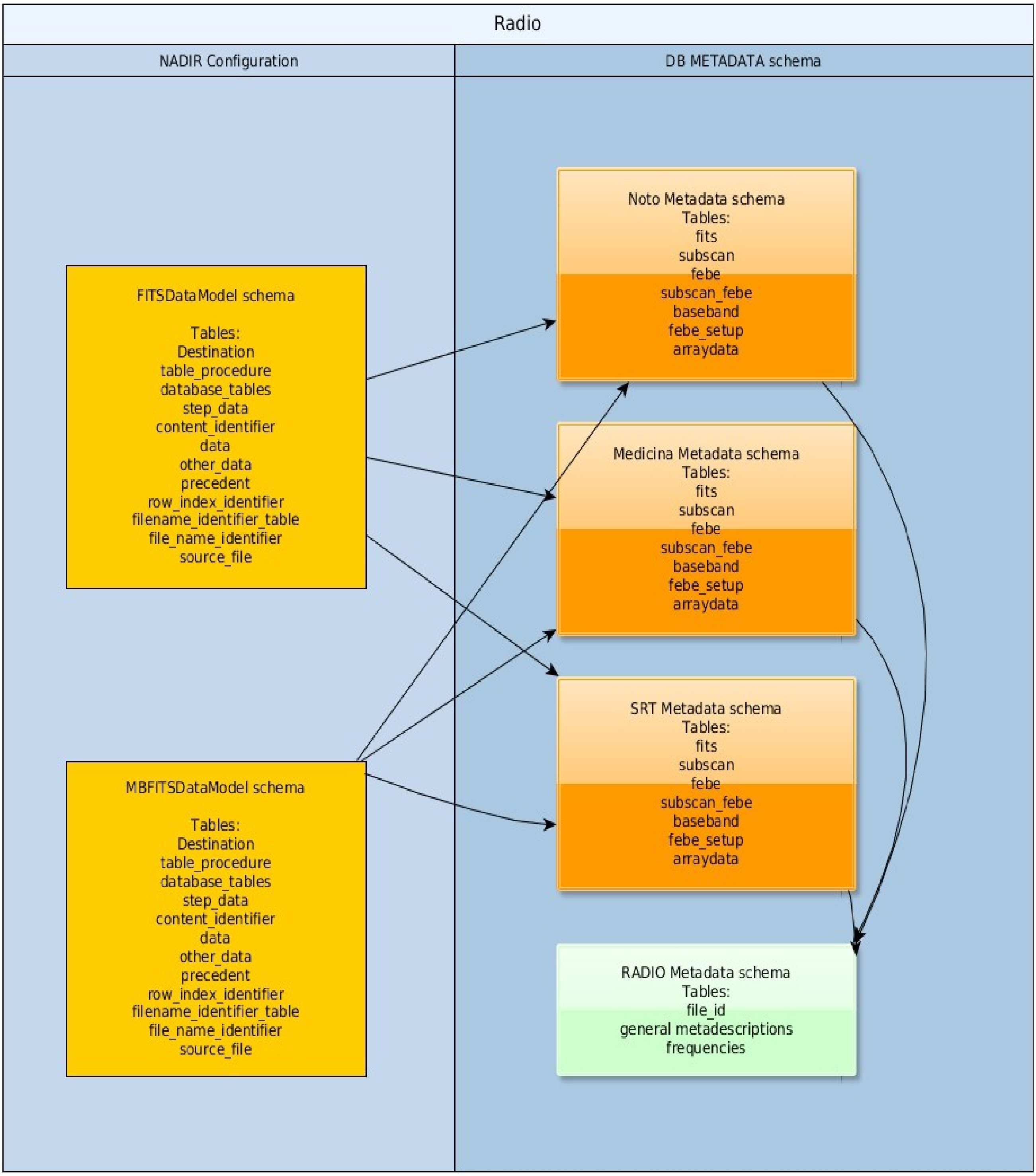


Figure E2: Entity relational diagram showing the software configuration model stored into database.

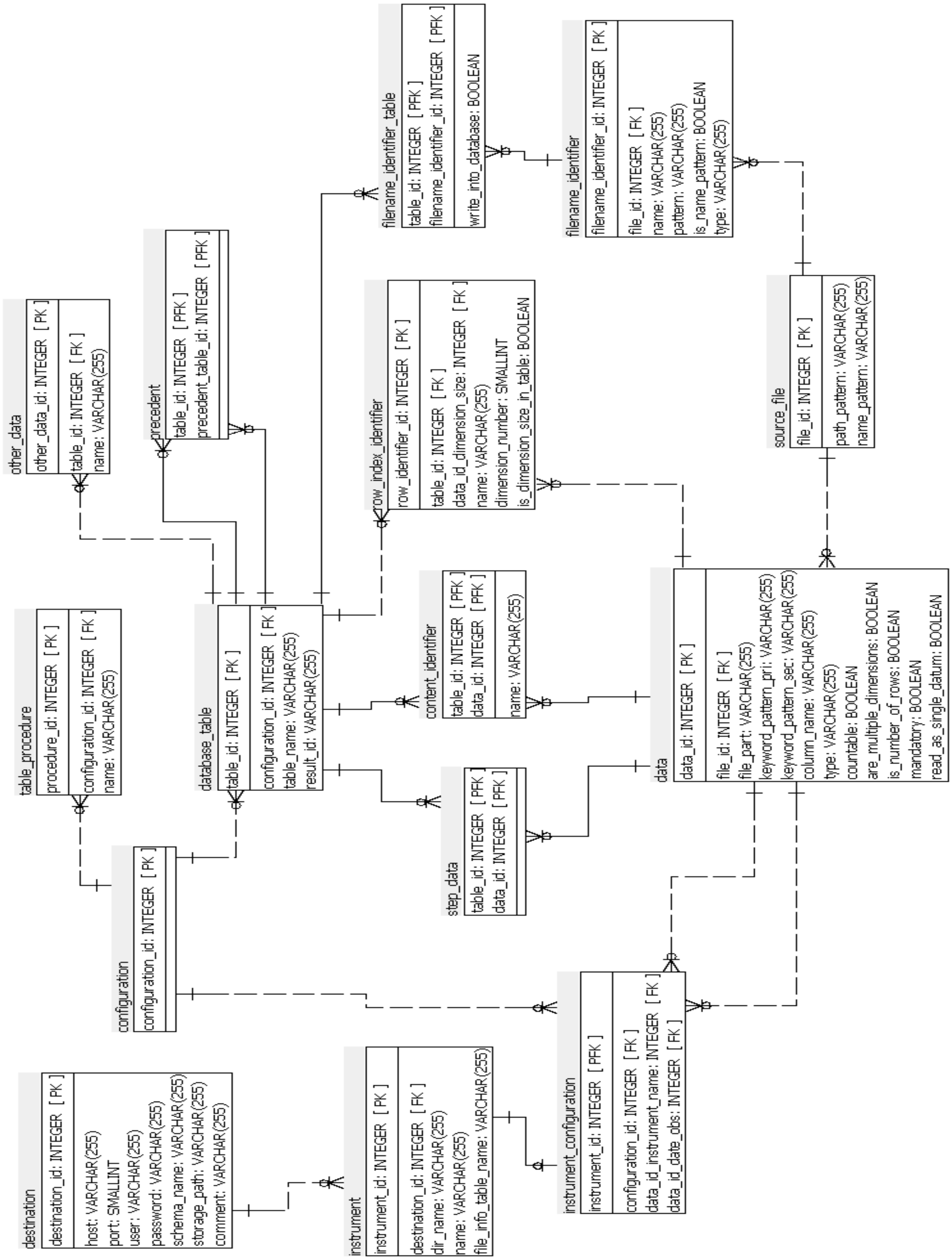
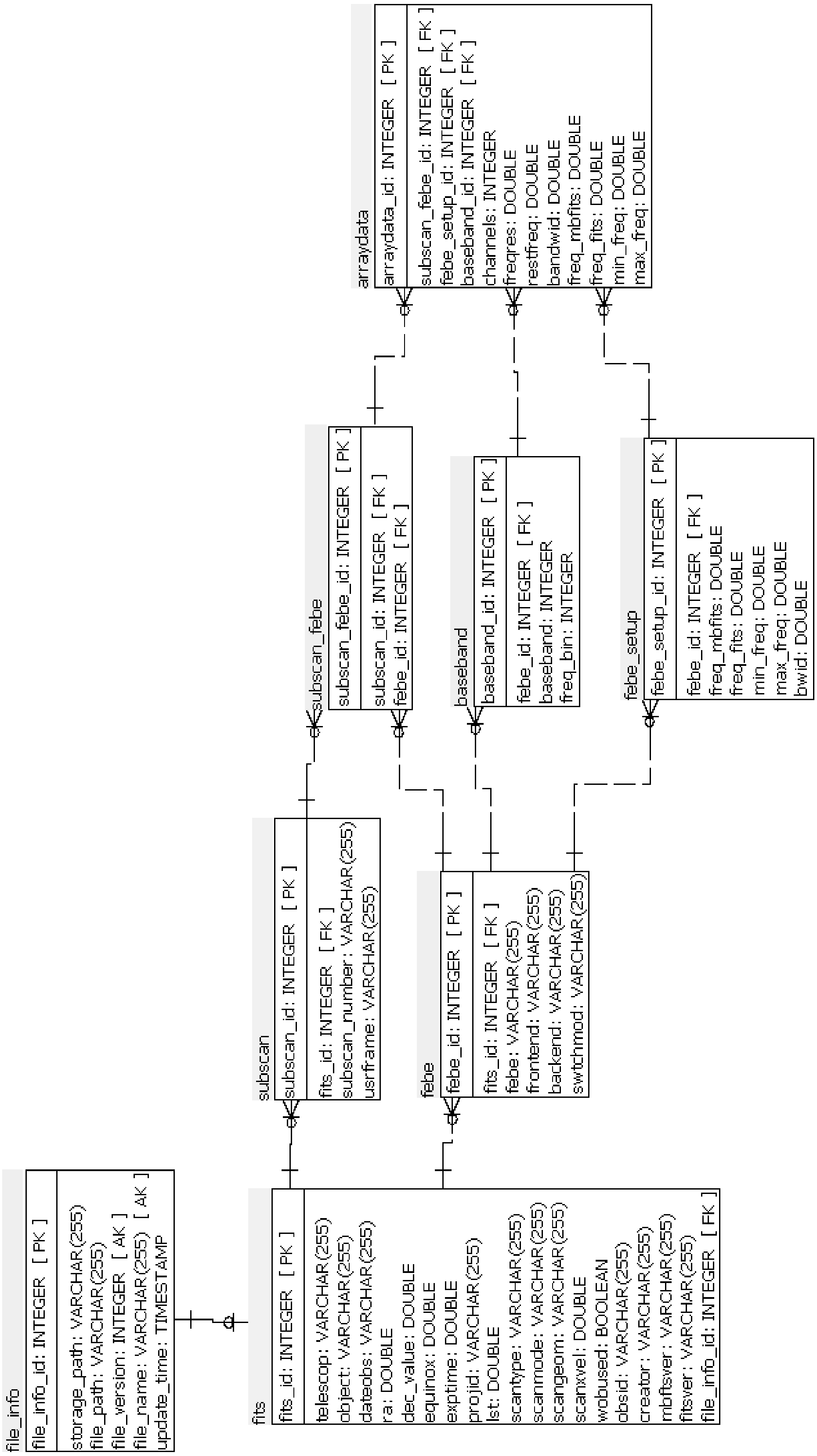


Figure E3: Entity relational diagram showing the data model of Radio MBFits and FITS files content.



REFERENCES

(1) *Multi-Beam FITS Raw Data Format Revision 1.65.* D. Muders, E. Polehampton and J. Hatchell, 2015. http://www3.mpifr-bonn.mpg.de/staff/dmuders/APEX/MBFITS/APEX-MPI-ICD-0002-R1_65.pdf.

(2) *A Hierarchical Grouping Convention for FITS, Rev. 8.* Jennings, D.G., Pence, W.D., Folk, M., Schlesinger, B.M., 1997, http://fits.gsfc.nasa.gov/registry/grouping.html