Radio Data Model for Medicina and Noto Telescopes

Cristina Knapic

EDP Forum and Training Event 2016 - Heidelberg

Outline

- IA2 Archives overview;

- Standards followed;

- Radio Data Formats;

- NEXT Data Model

- NEXT Configurability

- Current Radio Archive status

- VO compliancy

Centro Italiano Archivi Astronomici (IA2) main goals

- Archiving systems
- safety,
- data curation and preservation,
- distribution over several geographical distributed sites,
- providing services and tools (TWiki, work-flow, etc..)
- data publication in the VO

  of *Astronomical Data*

IA2 manages data of several PROJECTS.
Mainly they come from:

- TELESCOPES (raw; INAF ground based; **Radio**)
- SURVEYS (raw and/or calibrated)
- SIMULATIONS (ITVO)

Heidelberg 15/06/2016

IA2 Projects

Project Name	Project Type	Data Type	Data Amount Archive	Data Amount DB	UI	VO	1yr User Access UI	1yr User Access VO
LBT ✓✓✓✓	Telescope	image / spectra	10TB	320MB	✓	SIA	2.5K	47k
TNG ✓✓✓✓	Telescope	image / spectra	3TB	1.2GB	✓	SIA	12k	55k
GAPS	Survey (TNG)	image / spectra	850MB		✓(TNG), TWiki, Yabi			
Asiago ✓✓✓ < 1yr	Telescopes	image / spectra	400GB	290MB	✓	(TAP)	700	(70)
PESSTO	Survey (NTT)	image / spectra	7GB	1MB	✓			
hosted services								
WGE SDSS redshifts	data mining	catalogue		8GB		SCS / TAP		11k / 170
Planck	early release	catalogue		8MB		SCS		49k
TIRGO	IR camera	image	(Arcetri)	100MB		SIA		13k
WINGS < 1yr	Survey	image / catalogue	26GB	700MB		SIA / SCS		11k / 10k
INAF-IAPS EPN < 1yr	dust particles (NASA)	catalogue		2MB		TAP		1k
ITVO	Theoretical Simulation	mixed	1TB	1MB	✓		70	

Under development:

Project Name	Project Type	Data Type	Data Amount	UI	VO	1yr User Access
RADIO	Array/antennas	Images/spectra		✓	✓	
SKA.TM.OBSMGT	Observing tools	Meta-data				

Heidelberg 15/06/2016

Standards followed

IA2 at the moment manage Astronomical Data mainly in **FITS**, **Hierarchical FITS** and **MBFITS** formats, Textual format for images and spectra and **GADGET2** for simulations. IA2 host also survey pipeline and related products and provides support to a survey dedicated TWiki.

- **IA2 archives follows the directives of OAIS (Open Archival Information System) standard**:
    ➜ data are logically split in **data descriptors** and **data content**.

- **IA2 as a service follows the IVOA standards directives** and expose several VO services and several User Interfaces VO compliant.

Heidelberg 15/06/2016

Radio operational modes and data formats

Operational modes

- Sigle Dish (**SD**):
  - Vast number of possible instrumental set up parameters;
  - Design and implementation built on top of data/metadata structure of MBFITS standard (defined for APEX)
  - Two formats handled:
    - **FITS** (with the support of a Summary.fits file)
    - **MBFITS**
  - Atomic unit (FITS or MBFITS) supplied with night logs and observing schedules (ancillary files)

- Very Large Base Interferometer -IT (**VLBI-IT**) :
  - Medicina, Noto and Sardinia Radio Telescope Italian telescopes involved plus VLBI network;
  - Custom format for the archival purposes defined ad hoc:
    - **XML** custom summary file with the main configuration parameters (subset of the previously mentioned MBFITS data model)
  - Atomic unit (Visibility FITS file + XML summary) supplied with night logs and observing schedules (ancillary files)

Merged together in a single Radio Archive !

Structured RADIO Data (1)

MBFITS is based on FITS data format but organize the data and metadata content in a different manner for allowing storing of multifeed receivers, multiple beam observing and multiple frontend/backend and array receivers combinations.

IRA FITS and MBFITS used formats are based on MBFITS Standard. **Hierarchy in _MBFITS_ structure**:

   i. Number of sub-scans (m);
   ii. Front End Back End (FEBE) configuration (**n**);
   iii. Base Band (k).

FEBE configuration number determines:

1) <FEBE-NAME>-FEBEPAR.fits number in root dir of MBFits;
2) <FEBE-NAME>-ARRAYDATA-<BASEBAND>.fits and
 <FEBE-NAME>-DATAPAR.fits number in the sub-scan dir.

Base Band number determines:

1)<FEBE-NAME>-ARRAYDATA-<BASEBAND>.fits number
Associated to the same FEBE.

```
/MBF-ROOT
|
|-> GROUPING.fits
|
|-> SCAN.fits
|
|-> <FEBE-NAME>-FEBEPAR.fits
|
|-> /1
|   | -> <FEBE-NAME>-ARRAYDATA-<1>.fits
|   |
|   | -> <FEBE-NAME>-ARRAYDATA-<k>.fits
|   |
|   | -> <FEBE-NAME>-DATAPAR.fits
|   | -> MONITOR.fits
...
|-> /m
|   | -> <FEBE-NAME>-ARRAYDATA-<1>.fits
|   |
|   | -> <FEBE-NAME>-ARRAYDATA-<k>.fits
|   |
|   | -> <FEBE-NAME>-DATAPAR.fits
|   | ->  MONITOR.fits
```

Heidelberg 15/06/2016

Structured RADIO Data (2)

Hierarchical grouping directory structure and composition:

- main dir name accordingly to OBSDATE and PROPID (scan level);
  - GROUPING.fits
  - SCAN.fits
  - m<FEBE-NAME>-FEBEPAR.fits

- mSubdir per subscan named accordingly to the subscan number:
  - MONITOR.fits;
  - ARRAYDATA for each FEBE combination:
    - k<FEBE-NAME>-ARRAYDATA.fits;
    - <FEBE-NAME>-DATAPAR.fits ;

Atomic unit: tar archive composed by MBFITS folder And ancillary files (night log plus schedule)

```
/MBF-ROOT
|
|-> GROUPING.fits
|
|-> SCAN.fits
|
|-> <FEBE-NAME>-FEBEPAR.fits
|
|-> /1
|   | -> <FEBE-NAME>-ARRAYDATA-<1>.fits
|   |
|   | -> <FEBE-NAME>-ARRAYDATA-<k>.fits
|   |
|   | -> <FEBE-NAME>-DATAPAR.fits
|   |
|   | -> MONITOR.fits
|
...
|-> /m
|   | -> <FEBE-NAME>-ARRAYDATA-<1>.fits
|   |
|   | -> <FEBE-NAME>-ARRAYDATA-<k>.fits
|   |
|   | -> <FEBE-NAME>-DATAPAR.fits
|   |
|   | -> MONITOR.fits
```

Structured RADIO Data (3)

FITS structure and composition:

- Atomic unit composed by one or more FITS files (one for each subscan);
- Common folder for each scan FITS files plus a Summary.fits file and ancillary files like night logs and observing schedule;
  - nDATE-UT-PROJID-OBJECT-SCAN-SUBSCAN.fits
  - Summary.fits
  - Ancillary (night log plus schedule)

- Multiple feeds and multiple combination of FEBE and frequency settings are store in Summary.fits file in primary header;
- Multiplicity of parameter are discovered querying special keywords; recursive interaction on special keys like frequency are performed.

```
/ROOT
|
|-> Summary.fits
|
|-> FITS<1>.fits
|
|
...
|-> FITS<k>.fits
|
|-> Night.log
|
|-> Schedule.txt
```

Structured RADIO Data (4)

VLBI - IT structure and composition:

- Atomic unit: tar archive composed by:
  - Visibility fits file,
  - VEX file,
  - night log,
  - Summary.xml

- Metadata are a subset of the previous datamodel:
  - source name, source RA and DEC, type of observation (Imaging, line, pulsar, etc.), date of the observation, time spent on-source, frequency, antennas participating to the array, data rate, project ID of the observations, PI name;
  - VLBI-IT data will contain many keyword sets (one for each of the observed sources in the VLBI-IT dataset)

```
/ROOT
|
|-> Summary.xml
|
|-> FITS.fits
|
|-> Night.log
|
|-> Schedule (VEX)
```

NEXT's paradigms

- Based on TANGO Distributed Control System;
- Radio data model stored in a dedicated DB (data_model)
- Use of specific software configuration based on Instrumental set up defined in data_model database
- Handling of different format data
- Scalable in number of data importer devices available
- Policy and versions revised easily, in a flexible manner;
- Capability to develop in the major Object Oriented Languages (C++, Java, Python)
- Strong logging and error handling;
- Data distribution of RDB using stored procedures and stored parameters

NEXT Mandatory Requirements:
- **INSTRUMENT;**
- **OBS DATE;**

NEXT functional requirements:
- **data_model configuration!**

NEXT non functional requirements:
- **Coherent filling of fits keyword values in terms of types and values consistencies to allow query efficiency;**

NADIR
Extension for
Radio (NEXT)

IMAGE

Administration Layer

File Importer
(abstract class)

Metadata management

Import data

Import
metadata

Specialized
File Importer

Raw data

File
Detecting

Resolve
data
location

Read data
from archive

Lookup for
new
metadata

Metadata
Exporter

Data
Exporter

Metadata
Importer

Data
Importer

Remote Archive/s
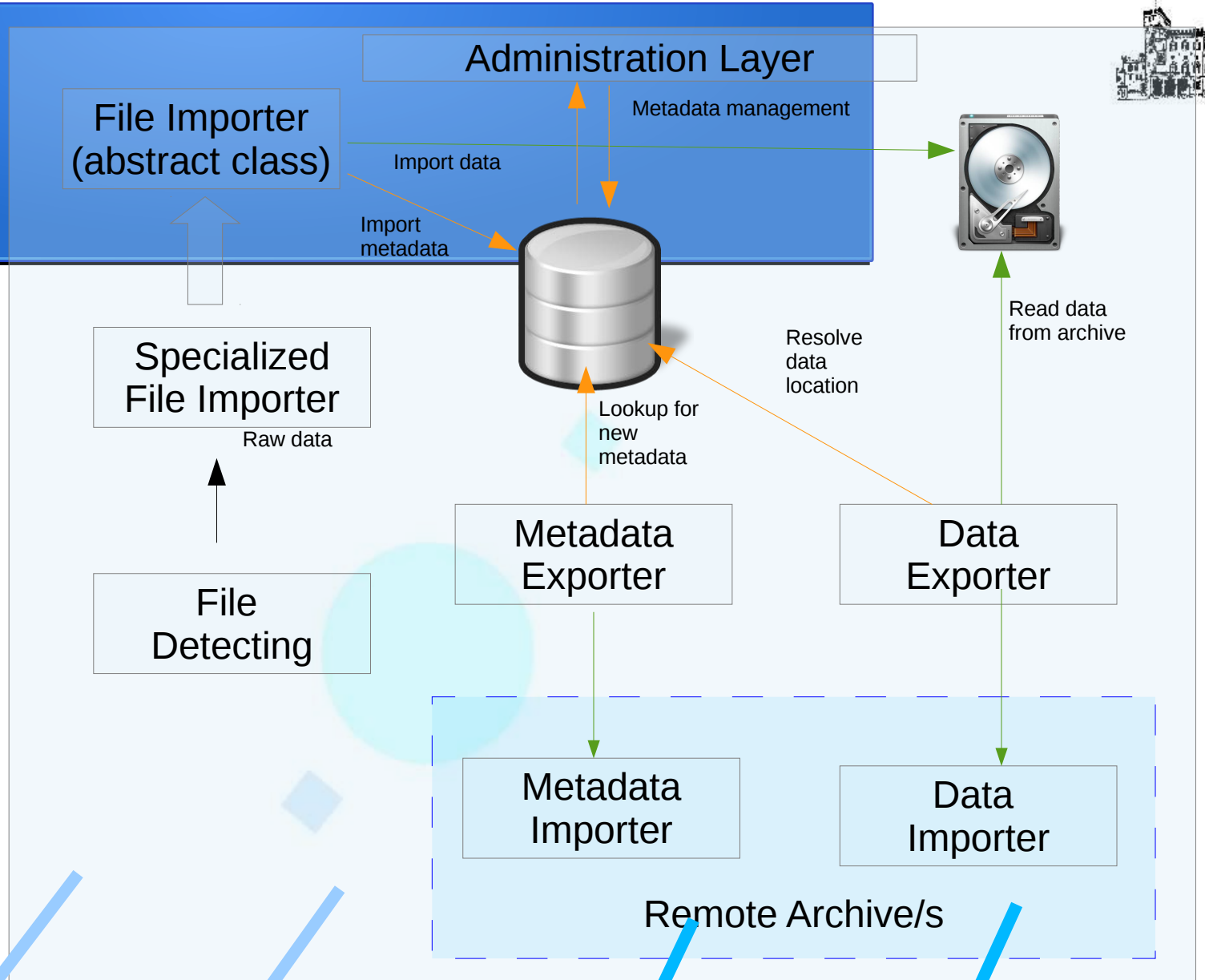
WEB INTERFACE

VO SERVICES

WEB INTERFACE

VO SERVICES

NEXT configuration database: keywords and identifiers

Keywords:
- a keyword can store several values (multidim. Arrays);
- A keyword can be defined with a pattern
  - i. Single (es. NFREQ)
  - ii. Countable (es. FREQn)
- A keyword can store the number of rows (es. NUSEBAND)
- A keyword can store a single value;

Identifiers: identify a table name or a table foreign key or a reference column/row;
- Content identifiers;
- Row index identifiers;
- file_name identifiers;
- Default identifiers.

NEXT configuration database: the structure and relation of data (1)

Data-model schema store information about:
- Instruments : each instrument has its own configuration setup;
  - Location of the Radio Archive DB and storages folders;
  - Structure of the Radio data Archive:
    - Name of tables;
    - Relations among tables;
    - Columns of each tables;
  - Description of the source files (naming conventions and data that can be found in each tar file content i.e. fits or xml).
- Configuration definition: definition of table and structure of the datamodel corresponding Radio Archive data base for each instrument and the source corresponding information (keyword location)
  - Definition of tables;
  - Definition of data in source files;
  - Data definition:
    - File where the data is stored;
    - Hierarchical point where finding the information (HDU number) or pattern;
    - Primary or secondary pattern;
    - Column name where store the data;
    - Type of data;

Heidelberg 15/06/2016

# NEXT configuration database: the structure and relation of data (2)

Configuration definition (cont.);
- Data definition (cont.):
  - Countable;
  - Are_multiple_dimension: i.e. reading HDU data content not metadata;
  - is_number_of_rows: i.e. CHANNELS store the number of rows related to channels table;
  - Mandatory;
  - read_as_a_single_datum: read bit by bit the value of this keyword to extract the right codification (library problem);
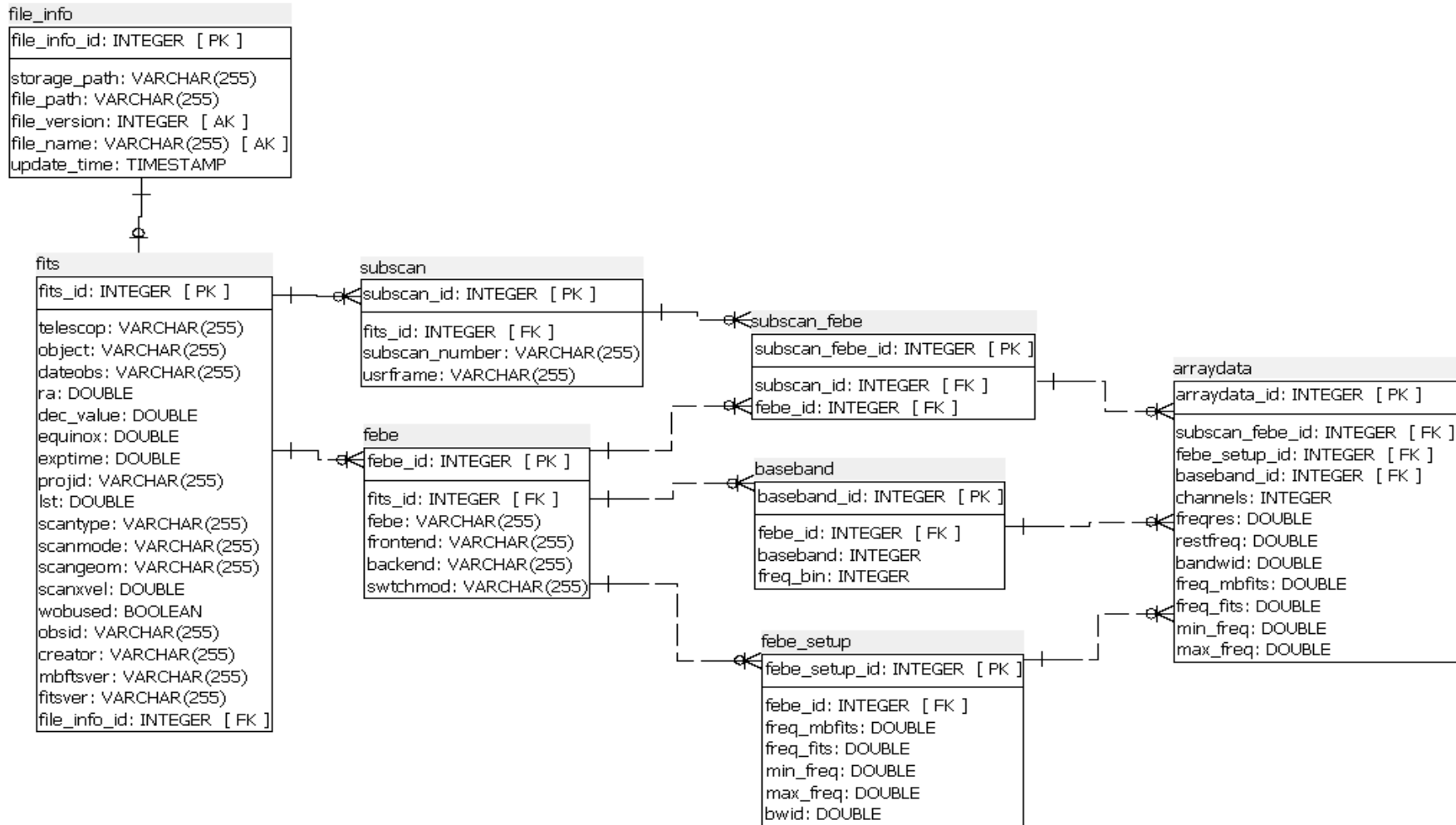- Definition of identifiers: foreign keys between entities;

Heidelberg 15/06/2016
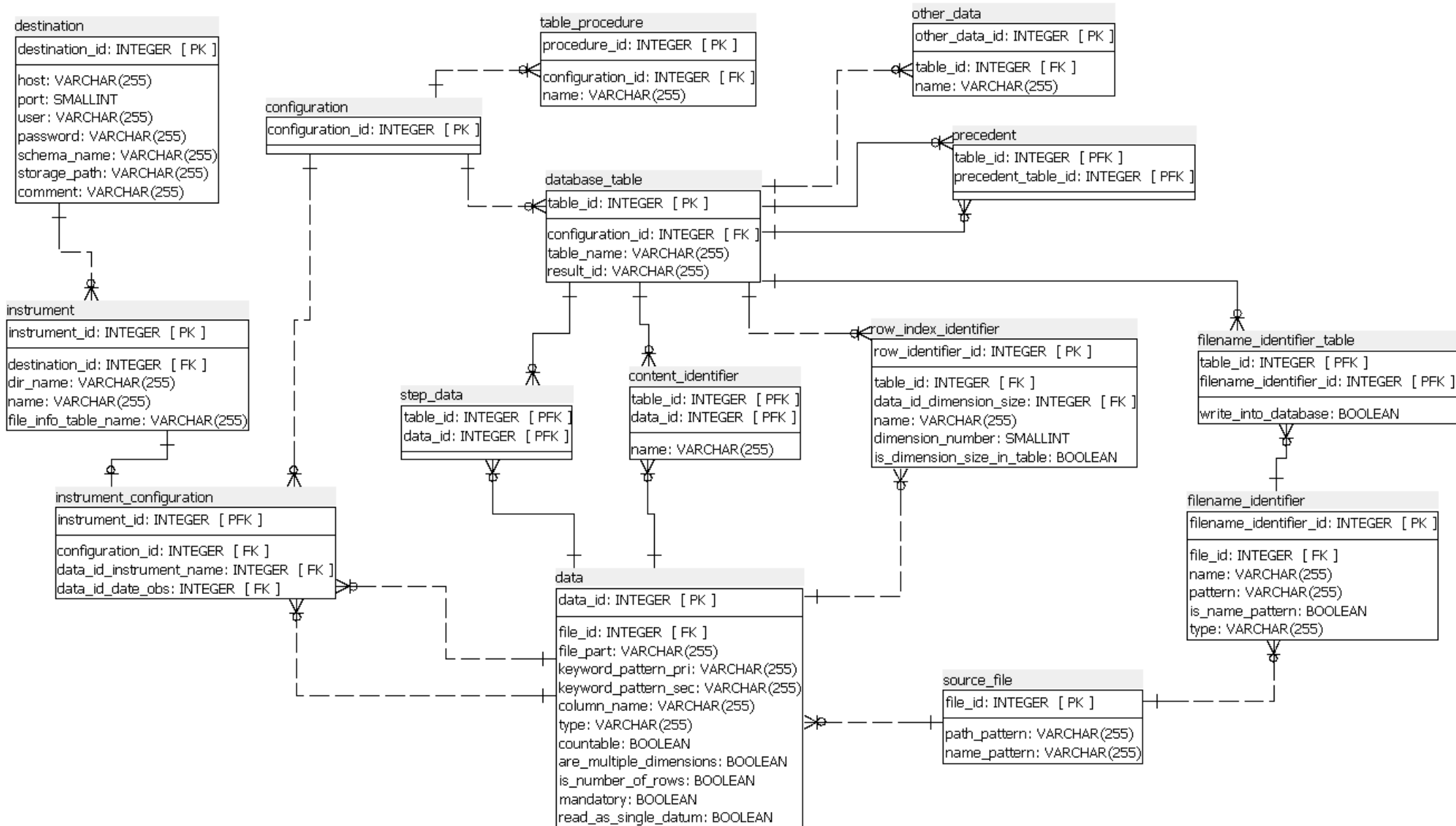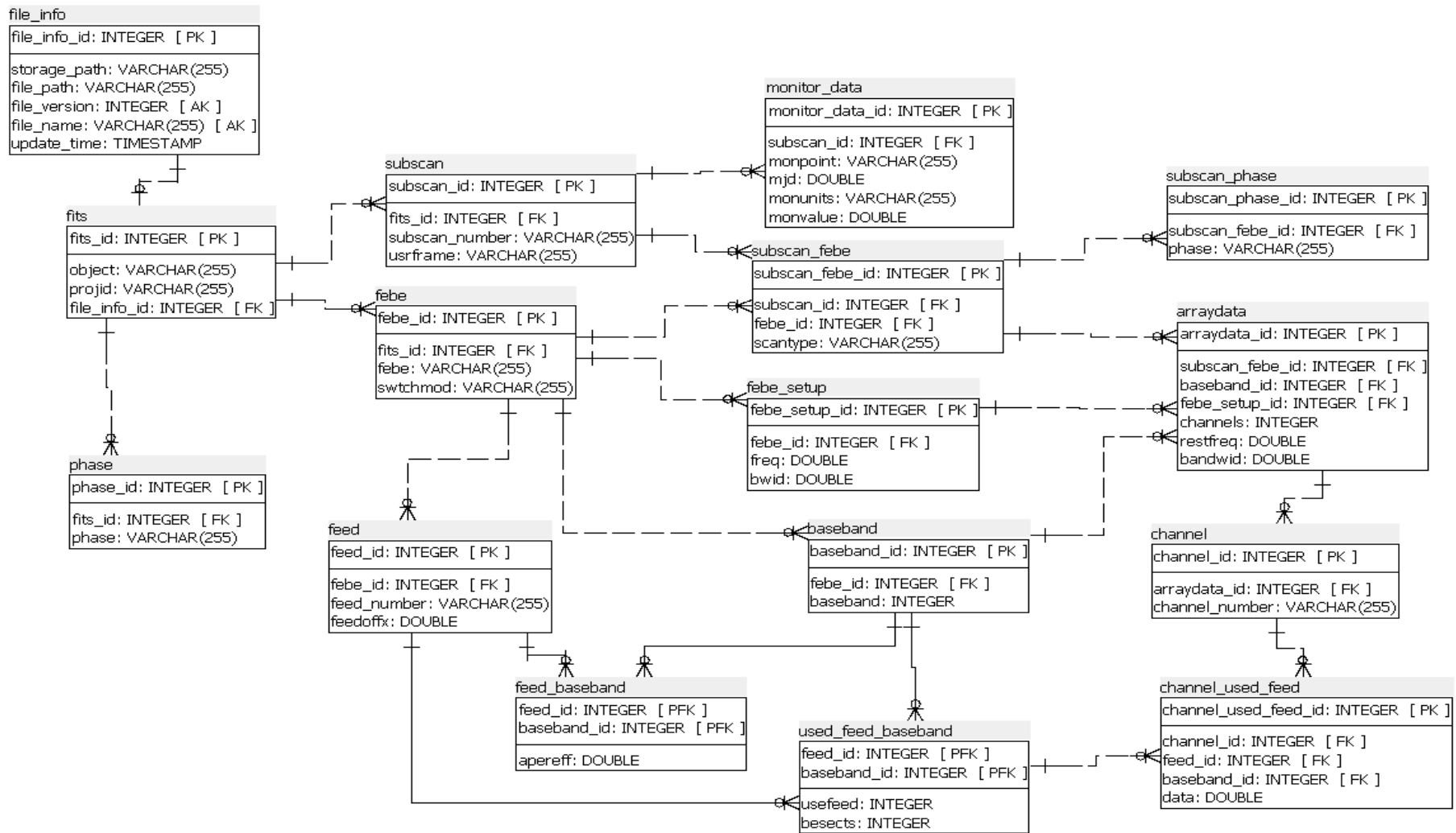
Radio Data Archive DB

Heidelberg 15/06/2016

Query optimization

DB indexing and store procedures helps to speed up the query execution.
Special SQL functions are installed in the DB and called by the external interfaces
to protect the DB from unoptimized usage.

VO compliance

Currently an investigation has started to understand if Radio Archive DB could be published using Table Access Protocol (TAP).
Open questions:
- Is the MBFITS Data Model fitting the specification of already existing models?
- Are the UCDS for all the parameters already defined?
- Could we integrate the SQL functions used to optimize query on DB into existing TAP standard implementations, instead of exporting fully or partially the radio schema?
- Is TAP the only service we can try to use to publish raw Radio data?
- Suggestions?

Thank you for your attention!!!

Cristina

in behalf of IA2 team

Heidelberg 15/06/2016