LOFAR Data Format ICD
TBB Time-Series Data

Document ID: LOFAR-USG-ICD-001

Version 3.04.01

SVN Repository Revision: 14203

L. Bähren, M. van den Akker, K. Anderson, A. Corstanje, A. Horneffer, J. Masters, L. Connor, S. ter Veen, G.A. Renting

SVN Date: November 19, 2018

Contents

List of Figures

List of Tables

Change record

0.1. Version 1.x

Notes for Version 1.0 → Version 1.1

1. *Reorganisation of the placement of attributes describing basic properties of the observation from which the dataset has originated.*
   Some part of the information which originally was supposed to be stored within the Station groups should be shifted upwards to the root group of the file.

   old: /
```
|-- STATION_001                         ... Group
|   |-- TELESCOPE                       ... Attribute       ... string
|   |-- OBSERVER                        ... Attribute       ... string
|   |-- PROJECT                         ... Attribute       ... string
|   |-- OBSERVATION_ID                  ... Attribute       ... string
|   |-- OBSERVATION_MODE                ... Attribute       ... string
|   `
|-- STATION_002                         ... Group
|   |-- TELESCOPE                       ... Attribute       ... string
|   |-- OBSERVER                        ... Attribute       ... string
|   |-- PROJECT                         ... Attribute       ... string
|   |-- OBSERVATION_ID                  ... Attribute       ... string
|   |-- OBSERVATION_MODE                ... Attribute       ... string
|   `
```
   new: /
```
|-- TELESCOPE                           ... Attribute       ... string
|-- OBSERVER                            ... Attribute       ... string
|-- PROJECT                             ... Attribute       ... string
|-- OBSERVATION_ID                      ... Attribute       ... string
|-- OBSERVATION_MODE                    ... Attribute       ... string
|-- STATION_001                         ... Group
|   |
|   `
|-- STATION_002                         ... Group
|   |
|   `
```

2. *Proper handling of coordinates*
   The motivation for this changes are in the fact, that for proper representation of a direction or position information more but just the numerical value is required – thus the additional information should be stored within the file as well, as compared to simply assuming the stored data adhere to a certain convention (which on the other hand they still should).

   old: /
```
|-- STATION_001                         ... Group
|   |-- BEAM_DIRECTION                  ... Attribute       ... array<double,1>
|   |-- 001000000                       ... Dataset
|   |   |-- ANTENNA_POSITION            ... Attribute       ... array<double,1>
|   |   |-- ANTENNA_ORIENTATION         ... Attribute       ... array<double,1>
|   |   |-- SAMPLE_FREQUENCY            ... Attribute       ... double
|
```
   new: /
```
|-- STATION_001                         ... Group
|   |-- STATION_POSITION_VALUE          ... Attribute       ... array<double,1>
|   |-- STATION_POSITION_UNIT           ... Attribute       ... string
|   |-- STATION_POSITION_FRAME          ... Attribute       ... string
|   |-- BEAM_DIRECTION_VALUE            ... Attribute       ... array<double,1>
```

```
    |   |-- BEAM_DIRECTION_UNIT            ... Attribute      ... string
    |   |-- BEAM_DIRECTION_FRAME           ... Attribute      ... string
    |   |-- 001000000                      ... Dataset
    |   |   |-- ANTENNA_POSITION_VALUE     ... Attribute      ... array<double,1>
    |   |   |-- ANTENNA_POSITION_UNIT      ... Attribute      ... string
    |   |   |-- ANTENNA_POSITION_FRAME     ... Attribute      ... string
    |   |   |-- ANTENNA_ORIENTATION_VALUE  ... Attribute      ... array<double,1>
    |   |   |-- ANTENNA_ORIENTATION_UNIT   ... Attribute      ... string
    |   |   |-- ANTENNA_ORIENTATION_FRAME  ... Attribute      ... string
    |   |   |-- SAMPLE_FREQUENCY_VALUE     ... Attribute      ... double
    |   |   |-- SAMPLE_FREQUENCY_UNIT      ... Attribute      ... string
    |
```

3. *Sensible (default) values*
   One of the – at least temporary – problems is, that some of the values to be stored within the HDF5 file are not available (yet) at the time of creating the file on disk. Therefore at least sensible default values/settings should be used to enable correct interpretation of the dataset; e.g. position information at a later point will be retrieved from the central parameter database, but until then the attributes should be filled already with placeholder values which agree with the conventions of the Measures framework. If a value in fact is undefined, it should be clearly marked as such by using `UNDEFINED` (in case of a string valued attribute).

```
/
|-- STATION_001
|   |-- STATION_POSITION_VALUE         ... array<double,1>    ...  {x,y,z}
|   |-- STATION_POSITION_UNIT          ... string             ...  "m"
|   |-- STATION_POSITION_FRAME         ... string             ...  "ITRF"
|   |-- BEAM_DIRECTION_VALUE           ... array<double,1>    ...  {0,90}
|   |-- BEAM_DIRECTION_UNIT            ... string             ...  "deg"
|   |-- BEAM_DIRECTION_STRING          ... string             ...  "UNDEFINED"
|   |-- 001000000
|   |   |-- ANTENNA_POSITION_VALUE     ... array<double,1>    ...  {x,y,z}
|   |   |-- ANTENNA_POSITION_UNIT      ... string             ...  "m"
|   |   |-- ANTENNA_POSITION_FRAME     ... string             ...  "ITRF"
|   |   |-- ANTENNA_ORIENTATION_VALUE  ... array<double,1>    ...  {x,y,z}
|   |   |-- ANTENNA_ORIENTATION_UNIT   ... string             ...  "m"
|   |   |-- ANTENNA_ORIENTATION_FRAME  ... string             ...  "ITRF"
|   |   |-- FEED                       ... string             ...  "UNDEFINED"
|   |   |-- NYQUIST_ZONE               ... uint               ...  1
|   |   |-- SAMPLE_FREQUENCY_UNIT      ... string             ...  "Hz"
|
```

Version 1.1 → Version 1.2

1. *Parametrization of coordinates*
   While the basic scheme for encoding and later reconstruction of the coordinate information remains unaltered, a minor detail was overlooked in the previous revision: in the general case one can not assume, that all values of a coodinate are given in the same physica unit. While e.g. for a simple direction – as described by two angles – the units are identical, this is not longer the when e.g. describing a position on the surface of the Earth (e.g. the position of the telescope). The simplest example for the latter case is the WGS84 system: in this a position is described by two angles and a height relative to the model geoid – therefore using `[deg,deg,m]` as units – hence the required modification in the data format to take this into account.

   old: /

```
    |-- STATION_001                        ... Group
    |   |-- STATION_POSITION_UNIT          ... Attribute      ... string
    |   |-- BEAM_DIRECTION_UNIT            ... Attribute      ... string
    |   |-- 001000000                      ... Dataset
```

```
    |    |    |-- ANTENNA_POSITION_UNIT       ... Attribute       ... string
    |    |    |-- ANTENNA_ORIENTATION_UNIT    ... Attribute       ... string
    |
```

new: /

```
    |-- STATION_001                           ... Group
    |    |-- STATION_POSITION_UNIT            ... Attribute       ... array<string,1>
    |    |-- BEAM_DIRECTION_UNIT              ... Attribute       ... array<string,1>
    |    |-- 001000000                        ... Dataset
    |    |    |-- ANTENNA_POSITION_UNIT       ... Attribute       ... array<string,1>
    |    |    |-- ANTENNA_ORIENTATION_UNIT    ... Attribute       ... array<string,1>
    |
```

As a consequence the values actually stored as attributes can look something like:

```
/
|-- STATION_001
|    |-- STATION_POSITION               ... array<double,1>    ...  {10,-6,50}
|    |-- STATION_POSITION_UNIT          ... array<string,1>    ...  {"m","deg","deg"}
|    |-- STATION_POSITION_FRAME         ... string             ...  "WGS84"
|    |
```

Version 1.2 → Version 1.x

`BEAM_WIDTH_VALUE`
`BEAM_WIDTH_UNIT`

Version	Date	Sections	Description of changes
1.00.00	yyyy-mm-dd	all	Attribute reorganization, "sensible default values".
1.01.00	yyyy-mm-dd	all	Coordinate parametrization.
1.02.00	2008-09-08	–	Beam parameters.
1.03.00	2008-09-08	–	–/–
1.04.00	2009-07-08	all	Reorganization wrt. common ICD format.
1.05.00	2009-09-15	3, 4.5	Updated high-level structure; new section on trigger table.
1.06.00	2010-01-26	1, 4, 6	Added summary of data volumes; update to trigger and calibration information; open questions now as table.
1.07.00	2010-02-03	4.5	Cleaning up of section on trigger table; extended list of table columns and description. Removed mentioning of Data Visualization Library (DVL).
1.08.00	2010-03-16	all	Merging various comments on the previous version of the ICD. Completely reworked section on station trigger.
1.09.00	2010-04-14	4.3, 6	Added attribute to station group. Added proposal for how to store calibration information as part of the station and dipole group; suggesting to replace simple dipole dataset by dipole group to take up both time-series data and calibration data.
1.10.00	2010-04-20	4/4.1	Refactor Root Group Sec. 4.1 → sec. 4.1.1, 4.1.2.
1.11.00	2010-05-14	3, 6.2	Integration suggested changes to dipole data structure into the main document.
1.12.00	2010-06-29	3.3,	Rewrite of section describing storage of calibration information. Added references.

Version 2.0 → Version 2.x

Version	Date	Sections	Description of changes
2.00.00	2010-07-08	Cover	Changed 'revision' to 'version'; updated this version number to 2.00.00 for LOFAR ICDs 1 through 7 to put them on the same version numbering scheme.

continued on next page

continued from previous page

Version	Date	Sections	Description of changes
2.01.00	2010-07-14	4.2, 4.1.2	Added *Acknowledgements* section. Adjusted version numbering scheme to include zero-padding. Added *clock offset* attribute. Adusted type of fit parameters in the *Station Trigger Group*.
2.01.01	2010-12-07	all	Using LaTeX package `hyperref` for references, enabling better navigation through the document and access to external resources.
2.01.02	2011-01-05	cover	the document Id is HARD. not svnInfo.
2.01.03	2011-02-17	all	Adding document title to page footer.
2.01.04	2011-02-28	all	Fixed data volume estimates, fixed naming conventions for Groups and Datasets to UpperCamelCase.
2.01.05	2011-03-10	all	Maintain list of references through BibLaTeX database.
2.01.06	2011-03-01	4	Added root group to hold trigger-specific information.
		Version upon which the current data-reader is based*	.
2.02.01	2011-04-20	all	Strings and string arrays: useage now consistent.
2.02.02	2011-05-12	all	Shifted all 2.01 and earlier changes to the'detailed change log' in the appendix.
		3	Added a 'data flow' placeholder, and switched the order of the 'overview' and 'heirarchical structure' subsections.
		4	Adjusted the 'coordinates subgroup' section in table 1.
		4	Removed the group-type for 'DumpMetaData' and replaced with 'string'.
		4	Added 'LORA' as a possible 'DUMP_TYPE' value.
		4	Ensured all Fields/Keywords are caps only.
		4	Changed dimensionality of `BEAM_SHAPE_DATA` to 3 (was 1). Everything is now 'BEAM_SHAPE'/'BeamShape'/'beam shape'.
		3 & 4	Station Trigger Group moved to the 'root' directory.
		all	Removed all but the first 'posix-style hierachy' from the document.
		3,4	Moved VHECR-specific component of 'station trigger group' to the root-level metadata.
2.02.07	2011-06-09	0,5, A	Re-organised appendices and included full change-log.
2.02.08	2011-07-06	all	Matching up group type attributes and notation; consolidation of labels to refer to standard sections and tables.
2.02.09	2011-09-15	all	Rework after review.
2.02.10	2011-12-21	0	Added data type section.
		4	Added metadata introduction.
2.02.11	2012-01-03	4	Added 'Manual' as keyword for trigger type.
2.02.12	2012-01-03	4	Update of optional keywords which are now in italic.
2.02.13	2012-01-11	cover	Modification of SVN version information.
		4	Change `CABLE_DELAY` keyword to `CABLE_DELAY_VALUE`.
2.02.14	2012-01-11	4	Added keyword `DIPOLE_CALIBRATION_DELAY`.
2.02.15	2012-01-31	4	Removed of calibration information.
			Added several keywords to the trigger group and dipole dataset.
			Updated mandatory/optional keywords.
2.02.16	2012-02-16		Renamed `ATTRIBUTE_VALUE` into `ATTRIBUTE`.

continued on next page

continued from previous page

Version	Date	Sections	Description of changes
			Renamed station group names: `STATION_NNN` → `AAKKK`
2.02.17	2012-02-21	all	Fixed all quotes for correct forward/backwards orientation.
2.02.18	2012-02-24	all	Added list of figures and list of tables + update of figure/table captions.
2.02.19	2012-03-05	5	Removed section 5 (Interfaces).
2.02.20	2012-03-06	cover	Added draftcopy package for background 'draft' text.
2.02.21	2012-03-07	all	Removed references to removed section 5 (Interfaces).
2.02.22	2012-03-09	4	Added column with software version number to keyword tables.
2.02.23	2012-03-20	4	Minor fix in Table 9.
2.02.24	2012-06-15	all	More consistent naming group names.
2.02.25	2012-07-12	all	Added minor modifications mentioned in ICD-meeting.
2.02.26	2012-07-19	4	Added definition of `SAMPLE_NUMBER` in text.
2.02.27	2012-11-13	4	Update of data structure description as they are now in the released DAL / TBB Writer.

*NOTE: P. Schellart used this document to design a data-reading program as of March/April 2011. Hence, everything from May onwards is designated as 'v2.02'.

Version 3.0 → Version 3.x

Version	Date	Sections	Description of changes
3.00.00	2017-11-07	all	Changed the document to a subband mode ICD. More cleanup is needed, but the relevant keywords are added.
3.01.00	2018-10-24	all	Cleanup of the document.
3.02.00	2018-10-24	all	Changed the document to support both raw voltage time series and subband mode time series.
3.03.00	2018-11-08	3,4	Clarified the overview text and re-organised the fields between Dipole Group and Subband dataset.
3.04.00	2018-11-16	3,4	Various small changes.
3.04.01	2018-11-22	4	Put DAL version 3.3 into the document and made some small typo fixes. Added glossary.

Version numbering scheme In order to track the evolution of the format specification documents the following numbering scheme has been adopted:

```
<major version>.<minor version>.<patch version>
  [0..]        .  [0..99]      .  [0..99]
```

where

- the `<patch version>` is getting incremented on changes to the document, which do not affect the actual contents of the file (such as when changing attribute names and such), e.g. correcting/augmenting descriptions, adding examples, etc.
- The `<minor version>` tracks minor changes to the actual content of the file, such as renaming, adding or removing attributes.
- The `<major version>` indicates major changes with in the file format, such as reorganization of the internal hierarchical structure or official release to the public.

Standard Data Types. The following table describes the short-form name for each type used throughout the rest of this document, it's logical meaning in the context of the astronomical data product, and the physical storage which must be allocated to it within the HDF5 data model. Future versions of this document may augment these types but will not remove support for existing types.

Name	Logical Type	Physical Storage Allocation
`short`	Integer; range -2^{15} to $2^{15}-1$	16 bit signed two's complement integer
`int`	Integer; range -2^{31} to $2^{31}-1$	32 bit signed two's complement integer
`unsigned`	Integer; range 0 to $2^{32}-1$	32 bit unsigned integer
`unsigned long long`	Integer; range 0 to $2^{64}-1$	64 bit unsigned integer
`float`	Single precision floating-point	IEEE 754-2008 [1] "binary32" floating point (1 bit sign, 8 bit exponent, 23 bit mantissa)
`double`	Double precision floating-point	IEEE 754-2008 "binary64" floating point (1 bit sign, 11 bit exponent, 52 bit mantissa)
`complex<type>`	Complex form of `type`	Compound type; the part at the lower memory location is real
`bool`	Boolean true/false	32 bit signed two's complement integer; non-zero denotes "true"
`string`	Text	Null-terminated string of 8 bit bytes. The lower 128 values interpreted as ASCII encoded characters
`array<type,N>`	Array of `type` with rank `N`	

Note that data may be written with either "big-endian" or "little-endian" byte ordering; either is valid within the context of this document.

Notation.

Symbol	Description
a, A	Italic lower and upper case chracters denote scalars.
$\mathbf{a}$	Bold lower case characters denote column vectors.
$\mathbf{A}_{[L,M]}$	Bold upper case characters denote matrices; (optional) if given $[L, M]$ denotes the shape.
a_i	Element i from vector $\mathbf{a}$.
A_{ij}	Element (i, j) from matrix $\mathbf{A}$.
$[name_0] \equiv$ `['Time']`	Array of rank 1, storing a single string-type value

1. Introduction

1.1. Purpose and scope

This interface control document (ICD) describes the internal structure of and the interface to the LOFAR time series data as generated by the Transient Buffer Boards (TBB). Time series data – i.e. the digitised voltage output, as received by the individual LOFAR dipoles – represent the primary input data to the CR (Cosmic Ray) analysis pipeline(s) and have to be considered as the most basic form in which the received radio signals are present within the LOFAR system.

1.2. Context and motivation

The fundamental difference between analysis for LOFAR Cosmic Ray (CR) data with respect to other LOFAR Key Science Projects (KSP) is the fact that processing starts from the raw digitised time-series data delivered by the individual dipoles of the LOFAR telescope. This approach is required to provide the necessary time-resolution – essentially down to the time-interval at which the analog signal is sampled – to detect, identify and investigate the radio pulses from Extensive Air-Showers (EAS) originating form high-energy cosmic rays.

More recently the raw digitised time-series data has also become useful in studying the science of lighting and thunderstorms, Fast Radio Bursts (FRB), and the interaction of neutrinos with the Moon.

Based on a number of considerations we have chosen the HDF5 data format as common wrapper for the standard LOFAR data products (or at least a considerable fraction thereof). The goal is to create along with the definitions of the standard data product also an infrastructure which will enable LOFAR users to access and manipulate such data – this document therefore also serves as reference for the implementation with the Data Access Library (DAL).

1.3. Applicable documents

Table 1 lists all the LOFAR ICDs. Most of the ICDs are for the various LOFAR data types, while ICD numbers 002 and 005 are general and applicable to all the data-format-oriented ICDs. Please note that the data and header information is written in Little-endian format within the HDF5 files.

Reference	Title	Description
ICD-001 [7]	TBB Time-Series Data	Digitized voltage output, as received by the individual LOFAR dipoles.
ICD-002 [6]	Representations of World Coordinates	Definition of how to represent and store meta-data that serve to locate a measurement in some multidimensional parameter space.
ICD-003 [2]	Beam-Formed Data	Hosting structure for LOFAR Beam-Formed data.
ICD-004 [8]	Radio Sky Image Cubes	Primary data product of the imaging pipeline.
ICD-005 [3]	File Naming Conventions	Conventions for the naming scheme applied to LOFAR standard data products.
ICD-006 [9]	Dynamic Spectrum Data	Hosting structure for dynamic spectrum data, i.e. intensity as function of time and frequency.
ICD-007 [5]	Visibility Data	Hosting structure for LOFAR UV Visibility data, primary output of interferometer operations.
ICD-008 [4]	RM Synthesis Cubes	Hosting structure for LOFAR Rotation Measure Synthesis Cubes output data.

Table 1: List of all the LOFAR Interface Control Documents. ICDs 001, 003, 004, 006, 007 and 008 describe different LOFAR data formats, while ICDs 002 and 005 are general and apply to the other ICDs.

2. Overview

This document is structured as follows: Section 3 will describe fundamental overall structure, including a statement of the primary data product format, HDF5. These conventions will also include names, meaning, and physical units that may be used to generate and interpret the data files. Section 4 will present a detailed specification for the data, including a description of the structure of a LOFAR TBB data naming conventions, units, physical quantities.

Comment:
This is not a finished document. It needs further clarification in the text of what certain values mean, especially in the Trigger Group section. See the various comment blocks in the text. The definition of the file format itself should be complete however, just not all values equally well documented.

3. Organisation of the data

3.1. High level LOFAR TBB Times-series file structure

A LOFAR TBB Times-series data file will adhere to the following guidelines:

A LOFAR TBB Times-series data file will be defined within the context of the HDF5 file format. A LOFAR TBB Time-series HDF5 file structure will comprise a primary group, a "root group" in HDF5 nomenclature, which may be considered equivalent to a primary header/data unit (HDU) of a standard multi-extension FITS file. This primary group will consist only of header keywords ("attributes" in HDF5 nomenclature) describing general properties of an observation, along with pointers to contained subgroups. Those subgroups will comprise an arbitrary number of "StationGroups" (see sec 4.2), where a Station Group will contain data and meta-data produced by an individual LOFAR station. Under each Station Group there are one or two more levels depending on the data taking mode used.

Figure 1 shows the basic organisation of the dataset within the HDF5 format. The hierarchical structure essentially follows the hierarchical structure of LOFAR itself, i.e. in a top-down approach from array through stations down to individual dipoles. The grouping of multiple antennas/dipoles into a station is mirrored by the collection of dipole datasets into a station group.

3.2. Overview of TBB Groups

From version 3 of this document, it is possible to store two variants of TBB data:

1. **Raw Voltage Mode** This is the same mode also supported by version 2 of this document. It records the raw voltage per dipole after analog-digital conversion. In older LOFAR documents and software this is often called *transient mode*. With a station clock of 200MHz this has a 5 ns resolution.

2. **Subband Mode** This variant is added in version 3 and allows to record each subband separately as complex voltages after the RSP has performed a polyphase filter (PPF) transpose into 512 subbands. The data for the subbands can be frozen with different time delays according to a specified dispersion measure (DM). In some older LOFAR documents and software this is called *spectral mode*. With a station clock of 200 Mhz this has a 5.12 µs resolution and up to 512 subbands.

The structure of the file is subdivided into the following four or five HDF5 group levels, depending on which mode is being used:

1. **File Root Group** (`ROOT`). The root level of the file contains the majority of associated meta-data, describing the circumstances of the observation. These data attributes include observation time (start and end), frequency window (high band vs. low band, filters) and other important characteristics of the dataset. See section 4.1 for further details.

2. **Trigger Group** (`TRIGGER`). This group collects parameters generated by the (station-level) trigger algorithm. See section 4.1.2 for further details.

3. **Station Group** (`STATION_{AAKKK}`). This group serves as a common container for the separate sub-tables, which take up data from the station calibration and the trigger algorithm. The part between `{...}` represents the station name and consists of 2 characters (upper case) and three digits (e.g. 'CS001' or 'RS106'). See section 4.2 for further details.

4. **Dipole Dataset/Group** (`DIPOLE_{NNNMMMLLL}`). This group collects data on a per-dipole basis starting from the identifiers required for the unambiguous identification of an individual dipole within the full LOFAR network to the actual sampled wave-form of the EM-field at the position of each antenna feed. This group either contains a dataset of the sampled raw voltages after analog-digital conversion or serves as a container for separate subtables for each subband in Subband Mode. The part between `{...}` represents the name of the dataset and is constructed from the *`STATION_ID`* (NNN), `RSP_ID` (MMM) and `RCU_ID` (LLL). See section 4.3 and 4.4 for further details.

5. **Subband Group** (`SB_{BBB}`). In Subband Mode this group contains the complex voltages for each subband. The part between `{...}` represents the the subband on the RCU and contains three digits (e.g. '100' or '106'). See section 4.5 for further details.

[**Comment:**
DIPOLE also doesn't seem to be the most accurate name, as we seem to be modelling HBA tiles or LBA dipoles?]

3.3. Hierarchical structure

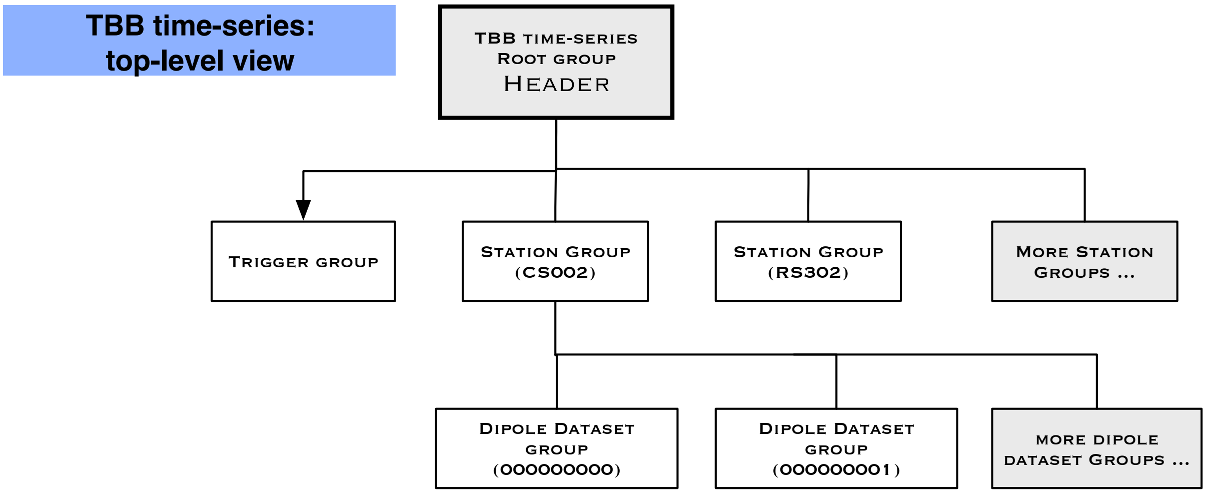

Figure 1: Hierarchical structure of a TBB times-series dataset; the internal organization of the data follows the hierarchical organization of the LOFAR system.

```
/                                     ...  Root Group
|-- TRIGGER                           ...  Group
|-- STATION_{AAKKK}                   ...  Group
|   |-- DIPOLE_{NNNMMMLLL}            ...  Dataset    [1D, short] (Raw Voltage Mode)
|   |-- DIPOLE_{NNNMMMLLL}            ...  Group      (Subband Mode)
|   |   |-- SB_{BBB}                  ...  Dataset    [1D, complex] (Subband Mode)
```

This structure can be represented through HDF5 as a POSIX-style hierarchy:
Raw Voltage Mode:

```
/
/TRIGGER
/STATION_CS002
/STATION_CS002/DIPOLE_000000000
/STATION_CS002/DIPOLE_000000001
/STATION_CS002/DIPOLE_000...
/STATION_RS302
/STATION_RS302/DIPOLE_001000000
/STATION_RS302/DIPOLE_001000001
/STATION_RS302/DIPOLE_001...
...
```

Subband Mode:

```
/
/TRIGGER
/STATION_CS002
/STATION_CS002/DIPOLE_000000000/SB_100
/STATION_CS002/DIPOLE_000000000/SB_101
/STATION_CS002/DIPOLE_000000000/SB...
/STATION_CS002/DIPOLE_000000001/SB_100
/STATION_CS002/DIPOLE_000000001/SB_101
/STATION_CS002/DIPOLE_000000001/SB...
/STATION_CS002/DIPOLE_000...
/STATION_RS302
/STATION_RS302/DIPOLE_001000000/SB_100
/STATION_RS302/DIPOLE_001000000/SB_101
/STATION_RS302/DIPOLE_001000000/SB...
...
```

4. Detailed Data Specification

LOFAR metadata are stored via Attributes in the HDF5 file. This section details the metadata within each HDF5 Group and Sub-Group. Attribute names, data types, default values, units and descriptions are summarized for each Group, in a table. The following conventions are used when storing data and meta-data in HDF5:

- Fields in *italics* are optional. All other fields are mandatory.
- All HDF5 groups have a `GROUPTYPE` attribute, that describes the type of the group. The ICD provides the mandatory value for this field for each group.
- `booleans` are stored as integers, with `0 = false` and `1 = true`. Implementations shall interpret any other value as `true` as well.
- Several meta-data come in (value, unit) pairs, such as `CLOCK_FREQUENCY` and `CLOCK_FREQUENCY_UNIT`. The `_UNIT` field describes the unit in which the meta-data is provided. Implementations shall:
  - Use the unit that is dictated in this ICD, unless doing so leads to an unacceptable loss of precision.
  - Use units that can be parsed by CasaCore.
- The following date and time formats are used:
  - `yyyy-mm-ddThh:mm:ss.sssssssssZ` for UTC timestamps, providing nanosecond precision.
  - `yyyy-mm-ddThh:mm:ss.sZ` for timestamps that need little precision, such as a file creation date.
  - `hh:mm:ss.sssssssss`, using an 24-hour clock, for UTC times.
- Missing data is represented by samples with a value of 0.

4.1. The Root Group

The LOFAR file hierarchy begins with the top level `'Root Group'`. This is the file entry point for the data, and the file node by which navigation of the data is provided. The `Root Group` will comprise a set of attributes that describe the underlying file structure, observational metadata, the LOFAR TBB data, as well as providing hooks to all groups attached to the `Root Group`.

This section will specify two sets of attributes that will appear in the `Root Group`: a set of Common LOFAR Attributes (CLA) that will be common to all LOFAR science data products, and a set of attributes that are specific to LOFAR TBB Time-series data. Though these attributes will all appear together in the `Root` attribute set, they are separated in this document in order to demarcate those general LOFAR attributes that are applicable across all data, and those attributes that are TBB-specific.

In other words,

```
Root Attributes = Common LOFAR Attributes (CLA) + Supplemental TBB Root Attributes.
```

The Common LOFAR Attributes are the first attributes of any LOFAR file root group.

4.1.1. Common LOFAR Attributes

This section will specify a set of attributes that will be common to LOFAR science data products. These "common LOFAR metadata" will appear as attributes at the root level of all LOFAR data files. *All* LOFAR data products, including TBB Time-series *inter alia,* will share a common set of metadata root-level attributes. These common LOFAR metadata are to be the first set of attributes of any LOFAR file root group.

Table 3 lists the Common LOFAR Attributes (CLA) which can be found in LOFAR observation mode data types: Beam-Formed, Transient Buffer Board (TBB) dumps, Time-series, and Sky Images within the files' root header. These Attributes are required to be in the Root Group; if a value is not available for an Attribute, a `'NULL'` maybe used in its place.

General LOFAR Group	Value	Description
Root	`'Root'`	Top-level LOFAR group type.

TBB Specific Groups	Value	Description
Trigger group	`'TriggerGroup'`	Trigger data group.
Station group	`'StationGroup'`	Station data group.
Dipole dataset	`'DipoleDataset'`	Individual Dipole Dataset (Raw Voltage Mode).
Dipole group	`'DipoleGroup'`	Individual Dipole Group (Subband Mode).
Subband dataset	`'SubbandDataset'`	Individual Subband Dataset (Subband Mode).

Table 2: LOFAR TBB Time-series Group Types.

4.1.2. Additional TBB Time-series Root Attributes

As explained at the beginning of Sec. 4.1 above, the root group of a **TBB Time-series data file** will contain a set of attributes, which can be broken down into two subsets: 1) a set of Common LOFAR Attributes (CLA) that will be common to all LOFAR science data products, and 2) a set of attributes that are specific to LOFAR Time-series data (see Table 7). With the Common LOFAR Attributes already listed in section 4.1.1 above, this section will focus on the second subset of root group attributes, as they are specific to a **TBB Times-series data file**.

Comment:
OPERATING_MODE seems to not be really well defined. "transient" = raw voltage mode? "spectral" = subband mode? Both are transient modes. Renaming things might break existing readers.

Field/Keyword	Type	Value	Software Version	Description
GROUPTYPE	string	'Root'	DAL v2.5.0	LOFAR Group type
FILENAME	string	—	DAL v2.5.0	File name
FILEDATE	string	—	DAL v2.5.0	File creation date
FILETYPE	string	—	DAL v2.5.0	File type (see Table 6)
TELESCOPE	string	'LOFAR'	DAL v2.5.0	Name of the telescope
PROJECT_ID	string	—	DAL v2.5.0	Project unique identifier
PROJECT_TITLE	string	—	DAL v2.5.0	Title of the project
PROJECT_PI	string	—	DAL v2.5.0	Name of Principal Investigator
PROJECT_CO_I	string	—	DAL v2.5.0	Name(s) of the Co-investigator(s)
PROJECT_CONTACT	string	—	DAL v2.5.0	Project contact details
OBSERVATION_ID	string	—	DAL v2.5.0	Observation unique identifier
OBSERVATION_START_UTC	string	—	DAL v2.5.0	Observation start date (UTC)
OBSERVATION_START_MJD	double	—	DAL v2.5.0	Observation start date (MJD)
OBSERVATION_END_UTC	string	—	DAL v2.5.0	Observation end date (UTC)
OBSERVATION_END_MJD	double	—	DAL v2.5.0	Observation end date (MJD)
OBSERVATION_NOF_STATIONS	unsigned	—	DAL v2.5.0	nof. stations used during the observation
OBSERVATION_STATIONS_LIST	array<string,1>	—	DAL v2.5.0	List of stations used during the observation
OBSERVATION_FREQUENCY_MIN	double	—	DAL v2.5.0	Observation minimum frequency
OBSERVATION_FREQUENCY_CENTER	double	—	DAL v2.5.0	Observation center frequency (weighted)
OBSERVATION_FREQUENCY_MAX	double	—	DAL v2.5.0	Observation maximum frequency
OBSERVATION_FREQUENCY_UNIT	string	'MHz'	DAL v2.5.0	
OBSERVATION_NOF_BITS_PER_SAMPLE	int	—	DAL v2.5.0	Number of bits per sample in the incoming data stream from the stations to CEP/BlueGene.
CLOCK_FREQUENCY	double	—	DAL v2.5.0	Clock frequency (160MHz or 200MHz).
CLOCK_FREQUENCY_UNIT	string	'MHz'	DAL v2.5.0	
ANTENNA_SET	string	—	DAL v2.5.0	Antenna set specification (see Table 4)
FILTER_SELECTION	string	—	DAL v2.5.0	Filter selection (see Table 5)
TARGETS	array<string,1>	—	DAL v2.5.0	List of observation targets/-sources
SYSTEM_VERSION	string	—	DAL v2.5.0	Processing system name/version
PIPELINE_NAME	string	—	DAL v2.5.0	Pipeline processing name
PIPELINE_VERSION	string	—	DAL v2.5.0	Pipeline processing version
DOC_NAME	string	—	DAL v2.5.0	Interface Control Document name
DOC_VERSION	string	—	DAL v2.5.0	Interface Control Document version/issue number
NOTES	string	—	DAL v2.5.0	Notes or comments

Table 3: Common LOFAR Attributes (CLA).

Antenna set	Description
`'LBA_INNER'`	48 antennas of the `INNER` LBA configuration.
`'LBA_OUTER'`	48 antennas of the `OUTER` LBA configuration.
`'LBA_SPARSE_EVEN'`	Intersection of `INNER-SPARSE` configurations.
`'LBA_SPARSE_ODD'`	Intersection of `OUTER-SPARSE` configurations.
`'LBA_X'`	X component, ALL LBA antennas.
`'LBA_Y'`	Y component, ALL LBA antennas.
`'HBA_ZERO'`	HBA antennas 0-23 in Core stations, and all HBA's in the other stations.
`'HBA_ONE'`	HBA antennas 24-47 in Core stations, and all HBA's in the other stations.
`'HBA_DUAL'`	Both HBA antenna (sub)fields in the Core stations, which set up an identical beam/pointing on each of those (sub)fields. On CEP, those (sub)fields are treated as seperate stations. On non-core stations, the whole HBA field is used and one beam is made.
`'HBA_JOINED'`	all HBA antennas in all stations types.
`'HBA_ZERO_INNER'`	Similar to HBA_ZERO, but with the inner half of the HBA antennas in use.
`'HBA_ONE_INNER'`	Similar to HBA_ONE, but with the inner half of the HBA antennas in use.
`'HBA_DUAL_INNER'`	Similar to HBA_DUAL, but with the inner half of the HBA antennas in use.
`'HBA_JOINED_INNER'`	Similar to HBA_JOINED, but with the inner half of the HBA antennas in use.

Table 4: Overview of antenna set configurations.

Filter-band, [MHz]	Attribute value
10 – 70	`'LBA_10_70'`
30 – 70	`'LBA_30_70'`
10 – 90	`'LBA_10_90'`
30 – 90	`'LBA_30_90'`
110 – 190	`'HBA_110_190'`
170 – 230	`'HBA_170_230'`
210 – 250	`'HBA_210_250'`

Table 5: Overview of filter-band selections and corresponding attribute values.

File Type	**Value**	**Description**
UV Vis	`'uv'`	LOFAR visibility file w/correlation UV information.
Sky cube	`'sky'`	LOFAR Image cube w/RA, Dec, frequency and polarization
RM cube	`'rm'`	Rotation Measure Synthesis Cube w/ axes of RA, Dec, Faraday Depth, polarization.
Near-field image	`'nfi'`	Near Field Sky Image w/ axes of position on the sky (x, y, z), frequency time, polarization.
Dynamic Spectra	`'dynspec'`	Dynamic Spectra w/ axes of time, frequency, polarization.
Beamformed data	`'bf'`	Beam-Formed file w/ time series data with axes of frequency vs time.
TBB dump	`'tbb'`	TBB dump file, raw time-series: (1) raw voltage mode: voltage vs time (baseband data) (2) subband mode: complex voltage vs time for selected frequency channels.
Instrument Model	`'inst'`	Parameters describing gain and other instrument characteristics for calibation.
Sky Model	`'lsm'`	List of sources, either point sources or shapelets.

Table 6: Overview of standard LOFAR data products and the corresponding file type attribute value.

Field/Keyword	Type	Description
`OPERATING_MODE`	`string`	Can either be "transient" or "spectral", The latter is not supported in DAL / TBB Writer v2.5(.0)
`NOF_STATIONS`	`unsigned`	The number of station groups available in the file.
`STATION_{AAKKK}`	Group	Station group collecting the data from an individual LOFAR station; the station name consists of 2 characters (upper case) and three digits.
`TRIGGER`	Group	Group collecting the parameters associated with the trigger and the output parameters generated by it.

Table 7: Additional attributes and objects attached to the root group of a TBB time-series data file.

Trigger Group The `TRIGGER` group collects the parameters that caused the TBB to be frozen and the data to be collected. There are three main types:

1. **External Trigger** These are triggers generated by a non-LOFAR instrument or telescope and then communicated to LOFAR. An example of this is a Fast Radio Burst detected by the WSRT of Effelsberg telescope and then communicated by VO Event.
2. **Internal Trigger** These are triggers generated by the (station-level) trigger algorithm, which can generate an "internal event" which will be responsible for causing the dump of the TBB data.
3. **LORA Trigger** These are triggers generated by the LORA cosmic ray detectors which also generate a sort of "inteternal event" on which the TBB can be triggered to freeze and dump data.
4. **Lightning Trigger** Something ??
5. **More Trigger** Something ??

Table 8 shows the attributes within this group.

Field/Keyword	Type	Software Version	Description
GROUPTYPE	string	DAL v2.0	The value of the grouptype is 'TriggerGroup'.
TRIGGER_TYPE	string	DAL v2.0	Type of trigger. The default value is "Unknown".
TRIGGER_VERSION	int	DAL v2.0	Version of the trigger algorithm.The default value is 0.
PARAM_COINCIDENCE_CHANNELS	int	DAL v2.0	The number of channels needed to detect a coincidence, or before it is an anti-coincidence. A typical value is 48.
PARAM_COINCIDENCE_TIME	double	DAL v2.0	The time-range in seconds, during which triggers are considered part of a coincidence. A typical value is 1e6
PARAM_DIRECTION_FIT	string	DAL v2.0	Do a direction fit.
PARAM_ELEVATION_MIN	double	DAL v2.0	Minimum elevation (in degrees) to accept a trigger. A typical value is 30.
PARAM_FIT_VARIANCE_MAX	double	DAL v2.0	Maximum variance ("badness of fit") of the direction fit to still accept a trigger. A typical value is 100.
COINCIDENCE_CHANNELS	int	—	Number of channels that took part in the coincidence.
COINCIDENCE_ID	array<int,1>	—	Identification
COINCIDENCE_FREQUENCY	array<double,2>	—	Frequencies used within the trigger channel.
COINCIDENCE_FREQUENCY_UNIT	string	—	Physical unit of COINCIDENCE_FREQUENCY.
COINCIDENCE_DIRECTION	array<double,1>	—	[2] Numercial value of the Tied Array beam direction.
COINCIDENCE_DIRECTION_UNIT	array<string,1>	—	Physical units associated with the numerical value of the Tied Array beam direction.
COINCIDENCE_DIRECTION_FRAME	string	—	Identifier for the reference frame within which the Tied Array-beam direction is provided.
TRIGGER_DISPERSION_MEASURE	double	DAL v3.3	Value of dispersion measure applied on the data before triggering.
TRIGGER_DISPERSION_MEASURE_UNIT	string	DAL v3.3	Unit of dispersion measure.
TIME	array<unsigned,1>	DAL v3.3	Timestamps in seconds since 1970.
SAMPLE_NUMBER	array<unsigned,1>	DAL v3.3	Sample numbers inside the second marked by TIME.
PULSE_SUM	array<double,1>	—	Sum of all the samples during the pulse.

Table 8: continued on next page

Table 8: continued from previous page

Field/Keyword	Type	Software Version	Description
PULSE_WIDTH	`array<double,1>`	—	Width of the pulse in samples.
PULSE_PEAK	`array<doube,1>`	—	The largest value (peak value) a sample had during the pulse.
PULSE_POWER_PRE	`array<double,1>`	—	Power before the onset of the pulse: value of the mean at the start of the trigger.
PULSE_POWER_POST	`array<double,1>`	—	Power before the onset of the pulse: value of the mean at the end of the trigger.
PULSE_RMS_PRE	`array<double,1>`	—	RMS value before the onset of the pulse: value of the RMS at the start of the trigger.
PULSE_RMS_POST	`array<double,1>`	—	RMS value before the onset of the pulse: value of the RMS at the end of the trigger.
THRESHOLD_LEVEL	`array<doube,1>`	—	The level above which triggers are excepted.
NOF_MISSED_TRIGGERS	`array<int,1>`	—	Number of missed triggers (+1) since the last trigger for this channel.
FIT_DIRECTION_COORDINATE_SYSTEM	`string`	DAL v3.3	Coordinate system for the direction fit.
FIT_DIRECTION_ANGLE1	`double`	DAL v3.3	Direction fit result for the (Azimuth) angle.
FIT_DIRECTION_ANGLE2	`double`	DAL v3.3	Direction fit result for the (Elevation) angle.
FIT_DIRECTION_DISTANCE	`double`	DAL v3.3	Direction fit result for the distance of curvature.
FIT_DIRECTION_VARIANCE	`double`	DAL v3.3	Variance ("badness of fit") of the direction fit.
REFERENCE_FREQUENCY	`double`	DAL v3.3	Reference Frequency of the dispersion measure calculation.
OBSERVATORY_COORDINATES	`array<double,1>`	DAL v3.3	Observatory coordinates of the dispersion measure calculation.
OBSERVATORY_COORDINATES_COORDINATE_SYSTEM	`string`	DAL v3.3	[ITRF??].
TRIGGER_ID	`string`	DAL v3.3	VO event coincidence ID or other identifier.
ADDITIONAL_INFO	`string`	DAL v3.3	Free form text field.

Table 8: Attributes attached to a TRIGGER group (TriggerGroup).

The type of trigger is given by the keyword `TRIGGER_TYPE` and contains one of the keywords from Table 9.

Comment:
TRIGGER_TYPE, TRIGGER_VERSION, TRIGGER_DISPERSION_MEASURE_*UNIT*, TIME, SAMPLE_NUMBER, FIT_DIRECTION_*, REFERENCE_FREQUENCY, OBSERVATORY_COORDINATES, TRIGGER_ID will all need to be present for a VO event based FRB trigger, where the ADDITIONAL_INFO can then contain the VO event text itself.

[**Comment:**
It is still unclear to me, if the fields like PULSE_* have any value as they are not implemented in the DAL. COINCIDENCE_ID seems to overlap with TRIGGER_ID]

Value of TRIGGER_TYPE	Description
'Unknown'	Unknown/unrecognised reason for data return.
'VHECR'	Single-station VHECR trigger.
'VHECRMulti'	Multi-station VHECR trigger.
'LORA'	Trigger from the LORA particle detector.
'FRATS'	Fast Radio Transients trigger.
'UHEP'	Trigger from Ultra-High Energy Particle mode (a.k.a 'Nu-Moon').
'Lightning'	Triggered in order to capture a lightning strike.
'Manual'	Manually triggered.
'FRB_VO'	Triggered by FRB Virtual Observatory Event.

Table 9: Values of TRIGGER_TYPE.

The remainder of the attributes can be divided into three groups

1. *Trigger algorithm setup parameters.*
   – PARAM_COINCIDENCE_CHANNELS marks the number of channels needed to detect a coincidence. The actual number of antennas which were part in the coincidence then is recoded through COINCIDENCE_CHANNELS.
2. *Trigger algorithm output parameters.*
   – COINCIDENCE_CHANNELS is the number of channels/dipoles, that took part in the coincidence; this number will be equal or larger as PARAM_COINCIDENCE_CHANNELS.
   – COINCIDENCE_ID Identification Number of the channels that took part in the coincidence. In case of the VHECR mode, these are RCU numbers.
   – TIME holds a list of the timestamps in seconds since 1970, for the RCUs which have been taken part in the coincidence.
3. *Fit results* based on the output parameters of the trigger algorithm.
   – FIT_DIRECTION_AZIMUTH and FIT_DIRECTION_ELEVATION are the fit results for the direction of arrival

4.2. The Station Group

Given the different modes for cosmic ray observation, a single LOFAR station is the natural choice for a first grouping of time-series data from the individual dipoles; for that matter we consider the **station group** (Table 10) as a basic module within the data structure.[1] Creating a snapshot of multiple stations, or even the full LOFAR array, thus will result in a set of station groups – which in turn might be collected into another superstructure.

The following entries will be found in the station group (Table 10, p. 23):

– GROUPTYPE identifies the group as a StationGroup.

[1]Though from initial perception the described structure can be perceived as a table, the HDF5 internal data model is that of a group; in order to stick as closely as possible to the libraries naming conventions, we therefore use the name *group* instead of *table*.

Field/Keyword	H5Type	Type	Software Version	Description
GROUPTYPE	Attr	string	DAL v2.0	LOFAR group type, StationGroup.
STATION_NAME	Attr	string	DAL v2.0	The name of the station, consisting of two characters (upper case) and three digits, e.g. CS001 or RS201.
STATION_POSITION	Attr	array<double,1>	DAL v2.0	[3] Numerical value of the station position coordinates.
STATION_POSITION_UNIT	Attr	string	DAL v2.0	Physical units associated with the numerical values for the station position.
STATION_POSITION_FRAME	Attr	string	DAL v2.0	Identifier for the reference frame within which the station position is provided.
BEAM_DIRECTION	Attr	array<double,1>	DAL v2.0	[2] Numercial value of the station-beam direction.
BEAM_DIRECTION_UNIT	Attr	string	DAL v2.0	Physical units associated with the numerical value of the station-beam direction. Mandatory for HBA station data.
BEAM_DIRECTION_FRAME	Attr	string	DAL v2.0	Identifier for the reference frame within which the station-beam direction is provided. Mandatory for HBA station data.
CLOCK_OFFSET	Attr	double	DAL v2.0	(Relative) Station clock offset.
CLOCK_OFFSET_UNIT	Attr	string	DAL v2.0	Physical unit for the station clock offset.
NOF_DIPOLES	Attr	unsigned int	DAL v2.0	The number of dipoles, for which data are embedded within this group.
DIPOLE_{NNNMMMLLL}}	Dataset	array<short,1>	DAL v2.0	Dataset containing the actual raw samples read out from the transient buffer; the name of the dataset is constructed from the *STATION_ID* (NNN), RSP_ID (MMM) and RCU_ID (LLL)
or				
DIPOLE_{NNNMMMLLL}}	Group	—	DAL v3.3	Container group for subband datasets in Subband Mode; The name is constructed from the *STATION_ID* (NNN), RSP_ID (MMM) and RCU_ID (LLL)

Table 10: Fields in the station data group (StationGroup). The main purpose of this group of to serve as a common container for the separate sub-tables, which take up data from the TBB. Shapes of vector and matrices are given in [ ]-brackets in the description. See text for detailed explanation on the individual fields in the table.

- While an internal identifier for the station is provided through the station ID, for better diagnostics the actual name of the station will be required; therefore `STATION_NAME` will store the actual name of the station, e.g. `CS001`, `RS201` or `DE602`.
- The **position of the LOFAR station** is reconstructed from the three attributes
  - `STATION_POSITION` – numerical value of the station position coordinates
  - `STATION_POSITION_UNIT` – physical units associated with the numerical values for the station position
  - `STATION_POSITION_FRAME` – identifier for the reference frame within which the station position is provided
- The **direction of the Tile Beam** on top of which the observation potentially has been running in piggy-back mode:
  - `BEAM_DIRECTION` – numerical value of the Tile Beam direction. This is the same as the direction of the first Sub Array Pointing and Station Beam. [2]
  - `BEAM_DIRECTION_UNIT` – physical units associated with the numerical value of the Tile Beam direction
  - `BEAM_DIRECTION_FRAME` – identifier for the reference frame within which the Tile Beam direction is provided
- `NOF_DIPOLES` is a counter for the number of dipoles, for which data are embedded within this group.

Even though there exist multiple LOFAR observation modes for TBBs, all have in common a (multi-level) pulse-detection and trigger-generation algorithm; the control parameters of the trigger algorithms as well as its output, in case a trigger condition was derived, need to be stored.

[2] Please note that as of writing in 2018, the Tile Beam, SAP0 and the Station Beam will all have the same pointing/direction, but LOFAR might become more flexible in the future with LOFAR 2

4.3. Dipole Dataset

The **Dipole Dataset** (Table 13) collects data on a per-dipole basis[3] – starting from the identifiers required for the unambiguous identification of an individual dipole within the full LOFAR network to the actual sampled wave-form of the EM-field at the position of each antenna feed.

Field/Keyword	Type	Software Version	Description
`GROUPTYPE`	`string`	DAL v2.0	The type of this group, *DipoleDataset*.
`STATION_ID`	`unsigned`	DAL v2.0	Data source station identifier.
`RSP_ID`	`unsigned`	DAL v2.0	Data source RSP board identifier.
`RCU_ID`	`unsigned`	DAL v2.0	Data source RCU board identifier.
`SAMPLE_FREQUENCY`	`double`	DAL v2.0	Sample frequency in MHz of the RCU boards.
`SAMPLE_FREQUENCY_UNIT`	`string`	DAL v2.0	Physical units of the sample frequency.
`TIME`	`unsigned`	DAL v2.0	Time instance in seconds of the first sample in the payload.
`SAMPLE_NUMBER`	`unsigned`	DAL v2.0	Sample number of the first payload sample in current seconds interval in transient mode.
`SAMPLES_PER_FRAME`	`unsigned`	DAL v2.0	Total number of samples in the payload of the original TBB–RSP frame structure.
`DATA_LENGTH`	`unsigned long long`	DAL v2.0	The number of samples per dipole which actually stored into the data set; this might as well be different from the number of samples in a data frame.
`FLAG_OFFSETS`	`array<unsigned long long,1>`	DAL v2.0	[2] Offset consisting of start and exclusive end position of the data range.
`NYQUIST_ZONE`	`unsigned`	DAL v2.0	Nyquist zone in which the data are sampled.
`CABLE_DELAY`	`double`	DAL v2.0	Delay the length of the cable connected to the RCU adds to the signal path.
`CABLE_DELAY_UNIT`	`string`	DAL v2.0	Physical unit associated with `CABLE_DELAY`.
`DIPOLE_CALIBRATION_DELAY`	`double`	DAL v2.0	Remaining delay to be applied to calibrate dipole up to station level.

Table 13: continued on next page

[3] Please keep in mind here, that we clearly distinguish between *antenna* and *dipole/feed*: using the feed-based approach as underlying the Measurement-Equation, an antenna can consist of multiple feeds (or dipoles).

Table 13: continued from previous page

Field/Keyword	Type	Software Version	Description
DIPOLE_CALIBRATION_DELAY_UNIT	string	DAL v2.0	Physical unit associated with DIPOLE_CALIBRATION_DELAY.
DIPOLE_CALIBRATION_GAIN_CURVE	array<complex,1>	DAL v2.0	Complex electronic gain as function of frequency.
ANTENNA_POSITION	array<double,1>	DAL v2.0	[3] Antenna position w.r.t. the station center, $\mathbf{x} = (x_1, x_2, x_3)$.
ANTENNA_POSITION_UNIT	string	DAL v2.0	Physical units of the antenna position.
ANTENNA_POSITION_FRAME	string	DAL v2.0	Reference frame of the antenna position.
ANTENNA_NORMAL_VECTOR	array<double,1>	DAL v2.0	[3] Antenna normal vector as specified in AntennaFields.conf files. Used to convert ITRF coordinates to local frame.
ANTENNA_ROTATION_MATRIX	array<double,1>	DAL v2.0	[9] Antenna rotation matrix as specified in AntennaFields.conf files. Used to convert ITRF coordinates to local frame. It is stored in row-minor order.
TILE_BEAM	array<double,1>	DAL v2.0	Numerical value of the tile beam direction.
TILE_BEAM_UNIT	string	DAL v2.0	Physical units of the tile beam orientation.
TILE_BEAM_FRAME	string	DAL v2.0	Reference frame of the tile beam.
DISPERSION_MEASURE	double	DAL v2.0	Value used for coherent dedispersion of the data.
DISPERSION_MEASURE_UNIT	string	DAL v2.0	Unit of dispersion measure.

Table 11: Fields in the dipole dataset; each listed field corresponds to a column in the table, where the number of rows corresponds to the number of dipoles. The first set of values is adopted directly from the frame structure used for data transfer between TBB and RSP [10]. Although the TILE_XXX keywords are optional for LBA station data, they are mandatory for HBA station data. Shapes of vector and matrices are given in [ ]-brackets in the description.

Comment:
Add attribute to store values of the analog beamformer for the HBA tiles and the dipole positions in the tile? There is some stuff about it in comments in versions 3.01 and older. do we mark if it is LBA or HBA? There was also some FEEDTYPE in older versions commented out.

Comment:
Do all of these make sense? Dispersion measure is not the DM of the trigger! Add ADC2VOLTAGE here as well?

– STATION_ID, RSP_ID and RCU_ID are directly taken from the frame structure used in the communication between RSP and TBB [10]. The three identifiers – in combination with ANTENNA_SET in combination – allow for an unambiguous identification of an individual dipole within the LOFAR network; depending on the range of value of the individual numbers the unique ID may be constructed via e.g.

$$N_{\mathrm{ID}} = 10^4 \cdot N_{\mathrm{Station}} + 10^2 \cdot N_{\mathrm{RSP}} + N_{\mathrm{RCU}} \tag{1}$$

– The combination of the two fields `TIME` and `SAMPLE_NUMBER` gives an absolute time reference for the first sample in the `DATA` field. The `TIME` field gives a time offset in seconds from a certain start moment, where the LCU is completely free at choosing a time system, such as UNIX time[4]. The `SAMPLE_NUMBER` key contains the sample number of the first payload sample in current seconds interval. For a constant sampling frequency (`SAMPLE_FREQ`) the timing for the remaining set of samples can derived via [10]

$$t[n] = t_{\mathrm{TIME}} + \left(t_{\mathrm{SAMPLE\_NR}} + n\right) \cdot 1/\nu_{\mathrm{SAMPLE\_FREQ}} \tag{2}$$

where n is the index for a sample in the `DATA` vector.

– Each frame of data transferred between RSP and TBB has the same fixed length (`SAMPLES_PER_FRAME`), but a frame may hold any number of samples that will fit in the payload area of the frame. As typically the number of samples requested from the TBB will be larger but the number of samples fitting into a single frame, the resulting dataset will accumulate the contents from multiple frames.

– `DATA` stores the raw ADC output for an individual signal path/dipole, consisting of `DATA_LENGTH` samples for a single dump of TBB data; the length of this data vector will vary depending on the observation mode.

– the **position** of the receiving element within the station:

- `ANTENNA_POSITION` – numerical value of the antenna position coordinates
- `ANTENNA_POSITION_UNIT` – physical units associated with the numerical values for the antenna position
- `ANTENNA_POSITION_FRAME` – identifier for the reference frame within which the antenna position is provided
- `ANTENNA_ROTATION_MATRIX` – this is used to convert ITRF coordinates to a local frame. It is stored as a 1-dimensional array of 9 element in row-minor order.

[4]Referring to the UNIX time, this field would hold the number of seconds since 1970.

4.4. Dipole Group

For the Subband Mode there needs to be another level of grouping below the **station group** called the **dipole group** (Table 12) to group the Subband Datasets.

Field/Keyword	H5Type	Type	Software Version	Description
GROUPTYPE	Attr	string	DAL v3.3	LOFAR group type, DipoleGroup.
STATION_ID	Attr	uint	DAL v3.3	[3] Numerical value for the station taken directly from the frame structure used in communication between the RSP and TBB boards.
RSP_ID	Attr	uint	DAL v3.3	[3] Numerical value for the RSP board taken directly from the frame structure used in communication between the RSP and TBB boards.
RCU_ID	Attr	uint	DAL v3.3	[3] Numerical value for the RCU board taken directly from the frame structure used in communication between the RSP and TBB boards.
SAMPLE_FREQUENCY	Attr	double	DAL v3.3	Sample frequency in MHz of the RCU boards.
SAMPLE_FREQUENCY_UNIT	Attr	string	DAL v3.3	Physical units of the sample frequency.
NYQUIST_ZONE	Attr	unsigned	DAL v3.3	Nyquist zone in which the data are sampled.
ADC2VOLTAGE	Attr	double	DAL v3.3	Conversion factor from raw ADC sample values to voltages
CABLE_DELAY	Attr	double	DAL v3.3	Delay the length of the cable connected to the RCU adds to the signal path.
CABLE_DELAY_UNIT	Attr	string	DAL v3.3	Physical unit associated with CABLE_DELAY.
DIPOLE_CALIBRATION_DELAY	Attr	double	DAL v3.3	Remaining delay to be applied to calibrate dipole up to station level.
DIPOLE_CALIBRATION_DELAY_UNIT	Attr	string	DAL v3.3	Physical unit associated with DIPOLE_CALIBRATION_DELAY.
DIPOLE_CALIBRATION_GAIN_CURVE	Attr	array<complex,1>	DAL v3.3	Complex electronic gain as function of frequency.
ANTENNA_POSITION	Attr	array<double,1>	DAL v3.3	[3] Antenna position w.r.t. the station center, $\mathbf{x} = (x_1, x_2, x_3)$.

Table 8: continued on next page

Table 8: continued from previous page

Field/Keyword	H5Type	Type	Software Version	Description
ANTENNA_POSITION_UNIT	Attr	string	DAL v3.3	Physical units of the antenna position.
ANTENNA_POSITION_FRAME	Attr	string	DAL v3.3	Reference frame of the antenna position.
ANTENNA_NORMAL_VECTOR	Attr	array<double,1>	DAL v3.3	[3] Antenna normal vector as specified in AntennaFields.conf files. Used to convert ITRF coordinates to local frame.
ANTENNA_ROTATION_MATRIX	Attr	array<double,1>	DAL v3.3	[9] Antenna rotation matrix as specified in AntennaFields.conf files. Used to convert ITRF coordinates to local frame. It is stored in row-minor order.
TILE_BEAM	Attr	array<double,1>	DAL v3.3	Numerical value of the tile beam direction.
TILE_BEAM_UNIT	Attr	string	DAL v3.3	Physical units of the tile beam orientation.
TILE_BEAM_FRAME	Attr	string	DAL v3.3	Reference frame of the tile beam.
DISPERSION_MEASURE	Attr	double	DAL v3.3	Value used for coherent dedispersion of the data.
DISPERSION_MEASURE_UNIT	Attr	string	DAL v3.3	Unit of dispersion measure.
NOF_SUBBANDS	Attr	uint	DAL v3.3	Number of subbands contained in a spectrum
SUBBANDS	Attr	array<uint,1>	DAL v3.3	List of all subbands selected; Items are between and include 0 to 511; The total should always be 487 long with the current TBB-ALERT firmware [version ??]
SB_{BBB}	Dataset	array<complex,1>	DAL v3.3	Dataset containing the actual raw samples read out from the transient buffer per subband; the name of the dataset is constructed from the SUBBAND NUMBER (BBB)

Table 12: Fields in the station data group (DipoleGroup). The main purpose of this group of to serve as a common container for the separate sub-tables, which take up data from the TBB. Shapes of vector and matrices are given in [ ]-brackets in the description. See text for detailed explanation on the individual fields in the table.

The following entries will be found in the dipole group (Table 12, p. 29):

- GROUPTYPE identifies the group as a DipoleGroup.
- STATION_ID, RSP_ID and RCU_ID are directly taken from the frame structure used in the communication between RSP and TBB [10]. The three identifiers – in combination with ANTENNA_SET in combination – allow for an unambiguous identification of an individual dipole within the LOFAR network; depending on the range of value of the individual numbers the unique ID may be constructed

via e.g.

$$N_{\mathrm{ID}} = 10^4 \cdot N_{\mathrm{Station}} + 10^2 \cdot N_{\mathrm{RSP}} + N_{\mathrm{RCU}} \tag{3}$$

– Each frame of data transferred between RSP and TBB has the same fixed length (`SAMPLES_PER_FRAME`), but a frame may hold any number of samples that will fit in the payload area of the frame. As typically the number of samples requested from the TBB will be larger but the number of samples fitting into a single frame, the resulting dataset will accumulate the contents from multiple frames.

– `DATA` stores the raw ADC output for an individual signal path/dipole, consisting of `DATA_LENGTH` samples for a single dump of TBB data; the length of this data vector will vary depending on the observation mode.

– the **position** of the receiving element within the station:
  - `ANTENNA_POSITION` – numerical value of the antenna position coordinates
  - `ANTENNA_POSITION_UNIT` – physical units associated with the numerical values for the antenna position
  - `ANTENNA_POSITION_FRAME` – identifier for the reference frame within which the antenna position is provided
  - `ANTENNA_ROTATION_MATRIX` – this is used to convert ITRF coordinates to a local frame. It is stored as a 1-dimensional array of 9 element in row-minor order.
  - `NOF_SUBBANDS` is a counter for the number of subbands, for which data are embedded within this group.
  - `SUBBANDS` identifies the subbands stored.

4.5. Subband Dataset

The **Subband Dataset** collects data on a per-subband per dipole starting from the identifiers required for the unambiguous identification of an individual dipole within the full LOFAR network to the actual sampled wave-form of the EM-field at the position of each antenna feed.

Field/Keyword	Type	Software Version	Description
GROUPTYPE	string	DAL v3.3	The type of this group, *DipoleDataset*.
TIME	unsigned	DAL v3.3	Time instance in seconds of the first sample in the payload.
CENTRAL_FREQUENCY	double	DAL v3.3	The central frequency of this subband.
CENTRAL_FREQUENCY_UNIT	string	DAL v3.3	The unit of the central frequency. [Hz]
BANDWIDTH	double	DAL v3.3	The bandwidth of this subband.
BANDWIDTH_UNIT	string	DAL v3.3	The unit of the bandwidth. [Hz]
TIME_RESOLUTION	double	DAL v3.3	The time resoltution of this subband.
TIME_RESOLUTION_UNIT	string	DAL v3.3	The unit of the time resolution. [s]
BAND_NUMBER	unsigned	DAL v3.3	Subband number.
SLICE_NUMBER	unsigned	DAL v3.3	This number indicates the slice(i.e. a transformed block of 1024 samples) of the first payload sample
SAMPLES_PER_FRAME	unsigned	DAL v3.3	Total number of samples in the payload of the original TBB–RSP frame structure.
DATA_LENGTH	unsigned long long	DAL v3.3	The number of samples per dipole which actually stored into the data set; this might as well be different from the number of samples in a data frame.
FLAG_OFFSETS	array<unsigned long long,1>	DAL v3.3	[2] Offset consisting of start and exclusive end position of the data range.

Table 13: Fields in the subband dataset; each listed field corresponds to a column in the table, where the number of rows corresponds to the number of time samples. Shapes of vector and matrices are given in [ ]-brackets in the description.

Comment:
Central frequency and time resolution could also be in a separate Spectral Group table under the root group directly, but that requires a larger restructuring of the data format.

A. Discussion & open questions

A.1. Open questions

The following table presents an overview of (some of the) known open questions regarding the format definition:

Item	Description	Raised by
01	Are there modes forseen in which the total LOFAR array is being split up into sub-arrays operating in different modes? In such a case the *range of application* of some of the metadata keywords would change; in order not to shift keywords within the data structure we therefore will end up with redundant information, depending of the specific observation mode. The latter though will not pose a major problem, since this redundance will show up in non-datasize critical keywords, such e.g. `OBSERVATION_MODE` or `NYQUIST_ZONE`.	L. Baehren
02	How to handle multiple HBA tile beams per station? Perhaps via subgroups of the dipole group?	L. Baehren
03	Is there indeed a separate value available which describes the conversion from ADC counts to voltages, or is this part of the gain calibration? If the latter is the case, then how to get a voltage time-series?	L. Baehren
04	Currently the trigger meta-data has no defined format. In the future, there will be a specific format for each trigger mode, i.e. an 'UHEP' group etc. In the meantime, the current format is to be used as a long character string that can be used flexibly.	C. James

A.2. Future enhancements

Though the file format definition is not intended to undergo considerable changes once gaining release status, there will be future enhancements to reflect new insights and address noted short-comings.

1. Some additional metadata are required when the observation is done using the HBA [input thanks to Maaijke Mevius]:
   a) At station level the 4x4 numbers which give the relative position of the dipoles in a tile, given in HBADeltas.conf. We do not really need those for analysis, but if you want to simulate the data (or do some sort of selfcal) it might be good to have them.
   Since these numbers are the same for all tiles within a station, I think they can be stored in the station metadata.
   b) The pointing of the tile beam (2 angles) should be added. I believe it is possible to have different tiles pointing in different directions, thus it should be stored per antenna.
   c) By 'tiles', do we mean literal tiles, or tiles/dipoles? [CWJ]

B. Anticipated Data volumes

The data format needs to be able to handle data volumes as different as a CR event with 1ms worth of data from a few antennas only to a full dump of 1 second worth of data from all LOFAR antennas in a consistent and efficient way.

UHEP-mode:

Time series data from the formed beam	
2 pol $\times$ 2^{27} samples $\times$ 8 Bytes/sample:	2.0 GB/event
Raw time series data from individual dipoles	
77 stations $\times$ 48 antennae $\times$ 2 pol $\times$ 2^{17} samples $\times$ 2 Bytes/sample:	1.8 GB/event
Total:	3.8 GB/event

VHECR-mode:

Raw time series data from individual dipoles	
48 antennae $\times$ 2 pol $\times$ 2^{17} samples $\times$ 2 Bytes/sample:	25 MB/event

HECR-mode:

Similar, but now only one station is involved	
48 antennae $\times$ 2 pol $\times$ 2^{17} samples $\times$ 2 Bytes/sample:	25 MB/event

TS-mode:

Similar to VHECR but full raw data from TBB	
77 stations $\times$ 48 antennae $\times$ 2 pol $\times$ $2 \cdot 10^8$ samples $\times$ 2 Bytes/sample:	3.0 TB/event

The event rate is uncertain, and for VHECR is estimated at one triggered event per station per 10 minutes. With 32 stations active this would amount to 110 GB per day of observing time.

C. Requirements

C.1. Metadata

Metadata is the auxiliary data stored along with the time series data need to provide all the necessary information for automated processing of the data. This data can either be stored directly in the data set or it can be stored in an external database. In the latter case the data set must contain a pointer to the correct entry in that database (e.g. the antenna-id is needed to get the antenna position).

- DAQ mode, including Samplerate, Filters, etc.
- Timing information:
  - Trigger time relative to recorded data segement
  - Timing of the data streams relative to the trigger
  - Timing of the data streams relative to each other with sub-sample accuracy; This can be implicit, e.g. all data streams of a station start at the same time. Fields that can be in stored in an external database.
- List of RFI sources identified by the station calibration, including **direction**, **center frequency** and **peak strength**.
  - → What does this actually mean: the properties of the single channel containing the highest signal level or the parameters obtained from fitting e.g. a Gaussian to a segment of the spectrum?
- Dispersion measure of the ionosphere (at this point in time and space)
- Health information about the antennas

C.1.1. System monitoring and system health.

Information on the status of the various (hardware) components at a LOFAR station will be stored inside the PVSS database; a description of the datapoint-types and datapoints can be found in the `MAC/Deployment/data/PVSS` branch of the LOFAR code repository (see Table 14 for an excerpt).

In order to later store certain system health information along with the other data, parameters need to be subscribed to at the definition of the observation.

C.2. Visualization

Past experience with the software for the LOPES experiment has shown, that is is crucial to provide the user with a variety of way to graphically inspect the data. This not only includes visualization of the time-series/FFT/etc. data themselves, but also displaying the various data with the wider context of the experimental setup (e.g. geographical distribution of the antennas w.r.t. to the particle detector setup)

- Display of the standard data products (also see documentation of the LOPES–`Tools DataReader`): ADC, Voltage, FFT, Calibrated FFT, RFI-filtered FFT, Cross-Corr. Spectra, Visibilities
- Display of the (intermediate) data products generated from the input data, e.g. dynamic spectra, multi-dimensional skymaps
- Flags and weights associated with the data, e.g. filter curves, antenna gain curves, etc.
- Station layout, i.e. positions of the (selected/excluded) antennas
- antenna power level distribution over the area of the station/array (this is very similar to the type of event display as known from particle physics experiments)
- mapping of the (local) RFI via an (Azimuth,Frequency) plot centered on the position of a certain station

Database entry	Field	Format
CalCtlr	connected	unsigned int
	obsname	string
	antennaArray	string
	filter	string
	nyquistzone	int
	rcus	string
ObservationControl	claimPeriod	int
	preparePeriod	int
	startTime	string
	stopTime	string
	subbandList	string
	beamletList	string
	bandFilter	string
	nyquistzone	int
	antenneArray	string
	receiverList	string
	sampleClock	int
	measurementSet	string
	stationList	string
	inputNodeList	string
	BGLNodeList	string
	storageNodeList	string

Table 14: Excerpt from the list of entries into the PVSS database. The definitions of the datapointtypes and datapoints can be found in the `MAC/Development/data/PVSS` branch of the LOFAR code repository.

- geographical locations of identified RFI sources; such a plot should also indicate the frequency band of the RFI (via label or bar etc.)
- geographical distribution/location of antennas which generated a trigger signal, failed, etc.
  - → VR setting, combining geographical information (e.g. map of the Netherlands) with the location of localized CR-/TS-events
  - ⇒ in a cave-like setup we actually can perform a fly-through, which would be ideal for outreach purposes!
- cummulative geographical distribution of detected CR events
- total power per LOFAR station (geographically distributed)
- overlay of CR data with information from other sensors (e.g. weather radar images, temperatures, etc.)

A number of the before-mentioned displayes should be interactive, in the sense that the user should be able to perform data selection from the graphical display (e.g. by drawing a circle around the core of the CR air shower, thereby selecting the antennas included in the data analysis step).

References

[1] IEEE Standard for Floating-Point Arithmetic. *IEEE Std 754-2008*, pages 1–58, 2008.

[2] A. Alexov, K. Anderson, L. Bähren, J.-M. Grießmeier, J. Hessels, J. Masters, and B. Stappers. Beam-Formed Data. LOFAR Data Format ICD LOFAR-USG-ICD-003, ASTRON, 2011.

[3] A. Alexov, K. Anderson, L. Bähren, A. Gunst, H. Holties, and M. Wise. File naming conventions. LOFAR Data Format ICD LOFAR-USG-ICD-005, ASTRON, 2011.

[4] J. Anderson, L. Bähren, M. Bell, T. Riller, and K. Anderson. Rotation measure synthesis cubes. LOFAR Data Format ICD LOFAR-USG-ICD-008, ASTRON, 2011.

[5] K. Anderson. Visibility data. LOFAR Data Format ICD LOFAR-USG-ICD-007, ASTRON, 2011.

[6] L. Bähren, A. Alexov, K. Anderson, and J.-M. Grießmeier. Representations of World Coordinates. LOFAR Data Format ICD LOFAR-USG-ICD-002, ASTRON, 2011.

[7] L. Bähren, K. Anderson, A. Corstanje, A. Horneffer, and J. Masters. TBB Time-Series Data. LOFAR Data Format ICD LOFAR-USG-ICD-001, ASTRON, 2011.

[8] L. Bähren, K. Anderson, and J. Masters. Radio sky image cubes. LOFAR Data Format ICD LOFAR-USG-ICD-004, ASTRON, 2011.

[9] J.-M. Grießmeier, A. Alexov, K. Anderson, and L. Bähren. Dynamic spectrum data. LOFAR Data Format ICD LOFAR-USG-ICD-006, ASTRON, 2011.

[10] W. Poiesz. Tbb design document. Software Design Document LOFAR-ASTRON-SDD-047, ASTRON, 2006.

Glossary of terms

Az Azimuth.

AIPS++ The AIPS++ project was a project from the nineties supposed to replace the original Astronomical Information Processing System or classical AIPS. The ++ comes from it being mainly developed in C++. It's also known as AIPS 2. It evolved into CASA, casacore and casarest (see those entries).

BBS BlackBoard Selfcal, pipeline used for LOFAR imaging data.

Beam A beam is formed by combining all the SubArrayPointing, one for each station, which are looking in a particular direction. There may be more than one beam for each SubArrayPointing, and different types of beams are available.

BF Beam-Formed data (time series structure).

CASA The Common Astronomy Software Applications package. User software for radioastronomy developed out of the old AIPS++ project. The project is led by NRAO with contributions from ESO, CSIRO/ATNF, NAOJ and ASTRON. [**?**]

casacore The set of C++ libraries that form the basis of CASA and several other astronomical packages. It contains classes for storing and handling visibility and image data, RDBMS-like table system and handling coordinates. Mainly maintained by ASTRON and CSIRO/ATNF. [**?**]

casarest The libraries and tools from the old AIPS++ project that are not part of casacore or CASA but still in use.

CEP Central Processing facility.

Channel The subband data of a LOFAR observation may be passed through a second polyphase filter to obtain a large number of channels (i.e. to increase the spectral resolution).

CLA Common LOFAR attributes. Set of root-level attributes that are used and required as attributes in all LOFAR science data products. If a value is not available for an Attribute, `'NULL'` maybe used.

Co-I Co-investigators on an observation project under the leadership of the PI.

Data Interface Set of definitions that describe the contents and structure of data files.

Data Access Layer (DAL) A C++ library with Python bindings providing read/write functionality for HDF5 format files, as well as access to Measurement Sets.

Dec Declination.

DPPP Default Pre-Processing Pipeline, pipeline used for LOFAR imaging data.

EAS Extensive Air-Shower.

El Elevation.

FITS FITS (Flexible Image Transport System) is a digital file format used to store, transmit, and manipulate scientific and other images. FITS commonly used in astronomy.

HBA High Band Antenna.

HDFView Hierarchical Data Format Viewer; a Java software tool for viewing the HDF5 structure and data. [http://www.hdfgroup.org/hdf-java-html/hdfview/]

HDF5 Hierarchical Data Format, 5 [**?**]. A file format capable of accommodating large datasets that comprises two (2) primary types of objects: groups and datasets. Implements self-organisation and hierarchical structures within the file format itself, facilitating self-contained data administration. [**?**, **?**]

HDF5 group A grouping structure containing zero or more HDF5 objects, together with supporting meta-data.

HDF5 dataset A multidimensional array of data elements, together with supporting meta-data.

HDU Header-Data Unit Though typically used for FITS data descriptions, the term "HDU" can also be used more generically when discussing any data group that contains both data and a descriptive header.

Hypercube The hypercube is a generalization of a 3-cube to n dimensions, also called an n-cube or measure polytope. In data modelling a hypercube is a cube-like logical model in which all measurements are organized into a multidimensional space.

ICD Interface Control Document.

IVOA International Virtual Observatory Alliance.

KSP Key Science Project. One of several major observational and research projects defined by the LOFAR organization. These Key Science Projects are,

- Cosmic Magnetism in the Nearby Universe
- High Energy Cosmic Rays
- Epoch of Re-ionization
- Extragalactic Sky Surveys
- Transients - Pulsars, Jet Sources, Planets, Flare stars
- Solar Physics and Space Weather

LBA Low Band Antenna.

LOFAR The LOw Frequency ARray. LOFAR is a multipurpose sensor array; its main application is astronomy at low radio frequencies, but it also has geophysical and agricultural applications. [http://www.lofar.org/]

LOFAR Sky Image Standard LOFAR Image Cube. A LOFAR data product encompassing science data, associated meta-data, and associated calibration information, including a Local Sky Model (LSM) , and other ancillary meta groups that are defined in this document.

LSM/GSM The Local Sky Model/Global Sky Model. Sky Models are essentially catalogues of known real radio sources in the sky. A Local Sky Model for an observation is merely a subset of a Global Sky Model catalogue pertaining to that observation's relevant region of the sky.

LTA The Long Term Archive for LOFAR.

MJD Modified Julian Day. Derived from Julian Date (JD) by MJD = JD - 2400000.5. Starts from midnight rather than noon.

MS Measurement Set, a self-described, structured set of casacore tables comprising the data and meta-data of an observation. [**?**]

PI A Principal Investigator is the lead scientist resopnsible for a particular observation project.

RA Right Ascension.

RFI Radio Frequency Interference.

RM Rotation Measure.

RMSC The Rotation Measure synthesis cube is a data product which contains the output of LOFAR RM synthesis routines, namely the polarized emission as a function of Faraday depth. As with the Sky Image data files, all associated information is stored within an RMSC file.

RSP Remote Station Processing Board.

SIP Standard Imaging Pipeline or Submission Information Package within the context of the LTA.

Station Group of antennae separated from other groups. In it's current cofiguration, LOFAR has 48 stations.

SubArrayPointing This corresponds to the beam formed by the sum of all of the elements of a station. For any given observation there may be more than one SubArrayPointing, and they can be pointed at different locations.

Subband At the station level, LOFAR data are passed through a polyphase filter, producing subbands of either 156.250 kHz or 195.3125 kHz (depending on system settings).

TAI International Atomic Time (Temps Atomique International), atomic coordinate time standard.

TBB Transient Buffer Board.

TRAP Transients Pipeline.

USG LOFAR User Software Group.

UTC Coordinated Universal Time (UTC) is a time standard based on International Atomic Time (TAI) with leap seconds added at irregular intervals to compensate for the Earth's slowing rotation.

UV-Coverage A spatial frequency domain area that must be covered completely by observation in order to assure an optimal target image (Full UV- Coverage). During observation, the radio telescope turns with respect to its target, due to the earth rotation. A certain -instrument geometry dependent-rotation angle has to be covered in order to accomplish full coverage.

VHECR Very high-energy cosmic ray.

WCS World Coordinate Information (WCS). The FITS "World Coordinate System" (WCS) convention defines keywords and usage that provide for the description of astronomical coordinate systems in a FITS image header [**?**, **?**, **?**].