# LOFAR Data Format ICD
# Radio Sky Image Cubes

**Document ID: LOFAR-USG-ICD-004**

**Version 2.06.05**

**SVN Repository Revision: 8212**

L. Bähren, K. Anderson, J. Masters

SVN Date: 2011-07-28

## Contents

## Change record

| Version | Date | Sections | Description of changes |
|---|---|---|---|
| 0.0 | 2009-04-03 | All | Initial revision |
| 0.1 | 2009-05-07 | All | Further full revision, new format |
| 0.2 | 2009-06-04 | 1.2, 3, | Revision on A2 comments |
| 0.3 | 2009-06-17 | All, + 4.4 | Full revision, migration to LaTeX |
| 0.4 | 2009-06-23 | 4.4 | Revision of attributes |
| 0.5 | 2009-07-01 | All | Structural overhaul, all ICDs |
| 0.6 | 2009-07-20 | 3, 4 | Structural adjustment, Coordinate group. |
| 0.7 | 2009-07-27 | 3, 4 , App. | Document redisposition, Filename, Glossary appendix |
| 0.8 | 2009-09-22 | All, 3.2+ | Recast group naming conventions, consistency |
| 0.9 | 2009-09-23 | 4.4 | Added missing coordinate type, added comments |
| 0.10 | 2009-09-29 | 4 | Cleaning up of coordinate group attributes |
| 0.11 | 2009-12-12 | 4 | Moved coordinates group table to external file |
| 0.12 | 2010-01-08 | 4 | Refactor LOFAR group type table |
| 0.13 | 2010-04-19 | 4 | Refactor S. 4.1 ==> 4.1.1, 4.1.2 |
| 0.14 | 2010-04-28 | 4.1.1 | New HBA naming convention (change in CLA) |
| 0.15 | 2010-05-03 | 4.1.1 | New `ANTENNA_SET` values, vis-a-vis `HBA` modes. |
| 0.16 | 2010-06-04 | Appendix | Removed section “Coordinate group examples” from the appendix; detailed description and examples now can be found in `LOFAR-USG-ICD-002` [?] |
| 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. |
| 2.01.00 | 2010-10-20 | 4.6 | Rewrite of the source table definition, as this was never fleshed out properly; description now also in sync with LOFAR-USG-ICD-008 [?]. Tracking of open question in the appropriate section of the document. |
| 2.01.01 | 2010-12-06 | all | Using LaTeX package `hyperref` for references, enabling better navigation through the document and access to external resources. Added note on version numbering scheme. |
| 2.02.00 | 2011-01-05 | 3, 4 | Major cleanup, diagram revision. Reset `GROUPTYPE` value examples in coordinates to read as `<CoordType>Coord`, eg., `‘StokesCoord’`. This convention had not been observed in some locations. See further comment at 4.4 , ‘The Coordinates Group.’ |
| 2.02.10 | 2011-01-05 | 4.4 | revamp file group to match diagram, other tweaks. |
| 2.02.11 | 2011-01-25 | 4.3 | Adjustment of attribute names to be more consistent; added labels and captions to tables and figures. |
| 2.03.00 | 2011-02-01 | 2 | move and insert image table, descr. text |
| 2.03.01 | 2011-03-10 | all | Maintain list of references through BibLaTeX database. |
| 2.04.00 | 2011-03-30 | all | Storage of source model parameters; Update of figures; Trimmed down number of coordinate sections. |
| 2.05.00 | 2011-04-27 | 4.5 | Attributes describing shape of data array |
| 2.05.01 | 2011-05-11 | 4.4 | Added paragraph with notation conventions |
| 2.05.02 | 2011-05-24 | 1, 2 | Cleaning up text, adjusting references. |
| 2.05.03 | 2011-05-25 | all | Addressing some of the comments provided by J. Swinbank. |
| 2.06.00 | 2011-05-25 | all | Removing *Average Images Groups* from the data structures tree; this does not add anything new, but rather confuses the overall architecture of the data file. Removed section “Interfaces”, which was not being used anyway. |

*continued on next page*

*continued from previous page*

| Version | Date | Sections | Description of changes |
|---|---|---|---|
| 2.06.01 | 2011-05-26 | 4.1.1 | Removed figure with antenna array configurations. |
| 2.06.02 | 2011-06-09 | all | Started addressing some of the comments from common reading of the ICD by the Data Formats Group. |
| 2.06.03 | 2011-06-21 | all | Second round addressing some of the comments from common reading of the ICD by the Data Formats Group. |
| 2.06.04 | 2011-06-29 | all | Matching up group type attributes and notation. |
| 2.06.05 | 2011-07-06 | all | Matching up group type attributes and notation; consolidation of labels to refer to standard sections and tables. |

**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.

**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 $\mathbf{A}$. |
| array<T,N> | Array of rank N, consiting of entries of type T |
| $[name_0] \equiv$ ['Time'] | Array of rank 1, storing a single string-type value |

## Acknowledgements

The document was originally developed by K. Anderson in collaboration with J. Masters.

# 1. Introduction

## 1.1. Purpose and Scope

The natural interface between the standard pipelines and the science pipelines is a well-defined set of standard LOFAR data products. These products are the outputs of the pipelines which run on Central Processing facility (CEP) and are the natural starting points for the science pipelines.

We note that any or all of these standard data products may also be accompanied by various ancillary data or "metadata". Metadata might include information such as instrument health, configuration or performance, gain curves, flagging tables, etc. In some cases this information may simply take the form of pointers to entries in the calibration database via header keywords. Alternatively, metadata may also be packaged in the same file with the main data in the form of additional tables or extensions.

This Interface Control Document sets forth a formal data interface specification for the LOFAR data products refered to as "**Radio Sky Image Cube**" and is intended to be the formal interface control agreement between the LOFAR project, observers/users of LOFAR data products, and the eventual LOFAR science archive facility.

## 1.2. Context and Motivation

A LOFAR **Radio Sky Image Cube** will be the data hosting structure for sky image data produced by LOFAR, irrespective of their scientific purpose. It is one of the tasks of the LOFAR project to define and describe the structure of the LOFAR **Radio Sky Image Cube** format.

While image hypercubes are the primary data product of the imaging pipeline, they are accompanied by a number of by-products, such as flagging information, CLEAN components, local and global sky models (LSM, GSM), etc. In the more traditional approach, where all such products are stored and managed separately, a large amount of book-keeping is required to maintain consistency. For the LOFAR project, a **Radio Sky Image Cube** product will be defined within the context of the Hierarchical Data Format 5, or HDF5 [**?**]. HDF5 allows for storage, not only of the data, but for the associated and related meta-data describing the image cube contents, conditions of observations, etc.. As an "all-in-one" wrapper, the HDF5 format simplifies the management of what are expected to be very large datasets that formats such as FITS cannot pragmatically accommodate.

There has been much discussion of a putative need for LOFAR **Radio Sky Image Cube** headers to adhere to FITS-like header keywords. Though it is envisioned that the LOFAR project will provide observers and other users with FITS format image files upon request, it is not entirely necessary that HDF5 header keywords match FITS keyword conventions in a LOFAR Sky Image file itself. A format conversion layer can certainly be developed to provide rigorous transformation of LOFAR headers into more restricted FITS header keyword sets. However, development of such a layer would be simplified in the event that LOFAR Sky Image files make use of *de facto* FITS standard keywords as much as possible.

For the purposes of further discussion regarding Sky Image file adherence to FITS keyword standards, the *ESO Data Interface Control Document* [**?**], has been adopted as the FITS keyword model.

# 2. Overview

LOFAR imaging data will be presented in a number of LOFAR data formats, all of which will provide data arrays of differing dimensions, depending upon the respective observation. Sky Image Cubes, Rotation Measure Cubes, Near-Field Cosmic Ray Images, etc., all have different dimensions and coordinate types. Table 1 illustrates the various image data array dimensions that LOFAR imaging may produce.

Each data type is described in detail by an appropriate interface control document. This document pertains to, and describes only those data conforming to the LOFAR datatype "Radio Sky Image".

The remainder of this document is structured as follows: Sec. 3 will present a high-level view of the hierarchical structure of LOFAR data files, file form, and semantic conventions the interface will adhere to, 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. Sec. 4 will present the low-level specification for the data, including a description of the structure of LOFAR Sky Image

| IMAGE | ICD | QUANTITY | AXES | UNITS |
|---|---|---|---|---|
| TBB time-series | 001 / [?] | $I(t)$ | Time | s |
| BF data | 003 / [?] | $I(p, \nu, \mathrm{Dec}, \mathrm{RA})$ | Pol/Freq/Dir/Dir | .. /Hz/deg/deg |
| Sky image | 004 / [?] | $I(p, \nu, \mathrm{Dec}, \mathrm{RA})$ | Pol/Freq/Dir/Dir | .. /Hz/deg/deg |
| Dyn. Spectrum | 006 / [?] | $I(p, \nu, t)$ | Pol/Freq/Time | .. /Hz/s |
| RMSC | 008 / [?] | $DF(p, \mathrm{Dec}, \mathrm{RA}, \phi)$ | Pol/Dir./Dir./Faraday Depth | .. /deg/deg/rad $\mathrm{m}^{-2}$ |
| RM map | — | $RM(\mathrm{Dec}, \mathrm{RA})$ | Dir./Dir. | /deg/deg |
| CR image | — | $I(p, \nu, r, \mathrm{El}, \mathrm{Az})$ | Pol/Freq/Dist/Dir./Dir./ | .. /p/Hz/m/deg/deg |
| CR image | — | $I(p, t, \nu, \xi_3, \xi_2, \xi_1)$ | Pol/Time/Freq/Pos/Pos/Pos | .. /s/Hz/m/m/m |

Table 1: Overview of the various data arrays types, associated coordinates and dimensions. Where possible a reference for the data format specification is provided.

files, and the various group entities and sub-structures comprising these image data files, i.e. LOFAR group types, units, physical quantities.

# 3. Organization of the data

## 3.1. High level LOFAR Sky Image file structure

A LOFAR **Radio Sky Image Cube** (a hypercube image file) will be defined within the context of the the Hierarchical Data Format, version 5 (HDF5). In an effort to minimize the hierarchical depth of the file structure, a Sky Image file is designed to be as "flat" as possible, providing access to the necessary data without undue hierarchical tree crawling.

Therefore, the Sky Image 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. These subgroups will comprise an arbitray number of "Image groups" (see Sec. 4.3), where an Image group will contain data and meta-data for a single sub-band of an observation.

## 3.2. Overview of Sky Image Groups

A LOFAR **Radio Sky Image Cube** will then comprise `System Log Group` just below the root level which contains logs and parameter files which are relevant to the entire file. Additionally, just below the root level, the Sky Image file will contain an arbitrary, observation-dependent number of `Image Groups` containing a `Coordinates Group`, a `Source Group`, and a `Processing History Group`, which contains pertinent logs and parameter sets of the relevent image sub-band.

These main building blocks of the Sky Image HDF5 file are (also see Fig. 1 below):

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 time, frequency and other important characteristics of the dataset. See Sec. 4.1 for a detailed description.
2. **System Logs Group** (`SYS_LOG`). This is a catch-all envelop encapsulating information about all the system-wide steps of processing which are relevant to the entire observation, such as parameter sets and processing logs. See Sec. 4.2 for a detailed description.
3. **Image Groups** (`IMAGE_{NNN}`). Each observation sub-band image is stored as a seperate group within the file, containing its own set of four (4) sub-groups. Characteristics about each sub-band image are stored as Attributes in group headers. LOFAR imaging can produce up to 244 sub-band images, which will be the maximum number of image groups possible in a LOFAR Sky Image. Each Image group will contain one Data group, which will in turn contain one dataset as an ndarray, along with associated attributes. See Sec. 4.3 for a detailed description.

4. **Source Table** (`SOURCE_TABLE`). Each observation sub-band processing produces a file describing what is called the Local Sky Model (LSM). A **Source Table** will contain a table of the sources listed in a processing file, a `.skymodel` file. Attributes will describe the columnar data. See Sec. 4.6 for a detailed description.
5. **Coordinates Groups** (`COORDINATES`). Each `Image Group` contains one `Coordinates Group`, which stores the relevant world coordinate conversions.
6. **Processing History Groups** (`PROCESS_HISTORY`) can be found on the `Image Group` level. These are catch-all envelops encapsulating information about all the steps of processing, such as parameter sets and processing logs. See Sec. 4.7 for a detailed description.
7. **Data Group/Image Data arrays** (`DATA`). For each Image Group, the subband image data are stored as ndarrays in the respective Data group – it is at this 4th hierarchical depth where the image data reside.

## 3.3. Hierarchical structure of the Sky Image file

A LOFAR **Radio Sky Image Cube** will then comprise an arbitrary, observation-dependent number of these Image groups.

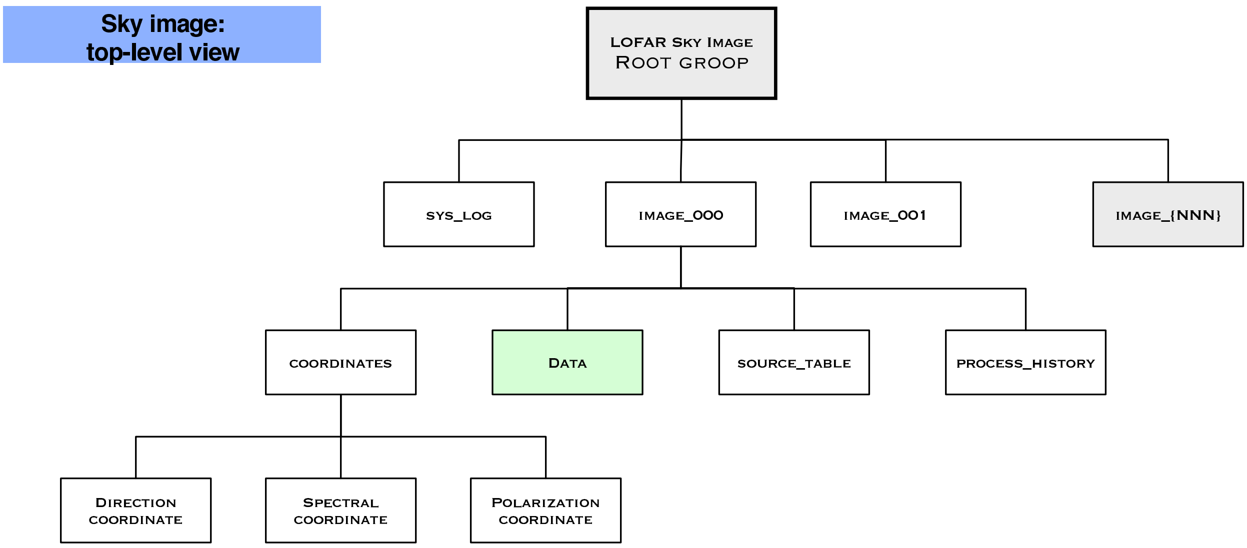


Figure 1: Top-level view on the structure of a sky image.

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

```
OBS_NUMBER/SYS_LOG
OBS_NUMBER/IMAGE_000
OBS_NUMBER/IMAGE_000/COORDINATES
OBS_NUMBER/IMAGE_000/COORDINATES/COORDINATE_0
OBS_NUMBER/IMAGE_000/COORDINATES/COORDINATE_1
OBS_NUMBER/IMAGE_000/COORDINATES/COORDINATE_2
OBS_NUMBER/IMAGE_000/IMAGE_DATA
OBS_NUMBER/IMAGE_000/SOURCE_TABLE
OBS_NUMBER/IMAGE_000/PROCESS_HISTORY
OBS_NUMBER/IMAGE_001
OBS_NUMBER/IMAGE_001/COORDINATES
OBS_NUMBER/IMAGE_001/COORDINATES/COORDINATE_0
OBS_NUMBER/IMAGE_001/COORDINATES/COORDINATE_1
OBS_NUMBER/IMAGE_001/COORDINATES/COORDINATE_2
OBS_NUMBER/IMAGE_001/IMAGE_DATA
OBS_NUMBER/IMAGE_001/SOURCE_TABLE
OBS_NUMBER/IMAGE_001/PROCESS_HISTORY
```

...

Similarly, one can represent the same structure as such:

```
L1002977_sky.h5                 ...   (Root) Group
|-- SYS_LOG                     ...   Group
|-- IMAGE_000                   ...   Group
|   |-- COORDINATES             ...   Group
|   |   |-- COORDINATE_0        ...   Group
|   |   |-- COORDINATE_1        ...   Group
|   |   `-- COORDINATE_2        ...   Group
|   |-- IMAGE_DATA              ...   Dataset
|   |-- SOURCE_TABLE            ...   Dataset
|   `-- PROCESS_HISTORY         ...   Group
|-- IMAGE_001                   ...   Group
|
`-- IMAGE{NNN}                  ...   Group
```

# 4. Detailed Data Specification

## 4.1. The Root Group (ROOT)

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

This section will specify two sets of attributes that will appear in the **File Root Group** of all LOFAR data product files: 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 Sky Image 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 image-specific. In other words,

`Root Attributes = Common LOFAR Attributes + Additional Root Group 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 Attributes" will appear as attributes at the root level of all LOFAR data files. *All* LOFAR data products, including Sky Images *inter alia,* will share a common set of metadata root-level attributes. These Common LOFAR Attributes are to be the first set of attributes of any LOFAR file root group.

Table 2 lists the Common LOFAR Attributes (CLA) which can be found in all LOFAR data file types. These attributes are required to be in the Root Group; if a value is not available for an attributes, a 'NULL' maybe used in its place.

– GROUPTYPE - The first Attribute in every group must be the attribute GROUPTYPE. Since the CLA are in the root header, the value in the CLA for (GROUPTYPE) = 'Root'. The options for the group type are listed in Tab. 6, grouped by category.

– FILENAME – Name of this file

– FILEDATE – File creation date, i.e. time at which the initial version of the file has been created.

– FILETYPE – is the **file type** for the LOFAR observation. This descriptor, which will also appear in LOFAR data filenames (see Table 3 below, or refer to [?]) of the LOFAR data file, indicates the kind of LOFAR data contained.

| Field/Keyword | Type | Value | Description |
|---|---|---|---|
| GROUPTYPE | string | 'Root' | LOFAR Group type (this is a 'root' group) |
| FILENAME | string | — | File name |
| FILEDATE | string | — | File creation date, i.e. time at which the initial version of the file has been created. YYYY-MM-DDThh:mm:ss.s |
| FILETYPE | string | — | File type |
| TELESCOPE | string | 'LOFAR' | Name of the telescope |
| OBSERVER | string | — | Name(s) of the observer(s) |
| PROJECT_ID | string | — | Unique identifier for the project |
| PROJECT_TITLE | string | — | Title of the project |
| PROJECT_PI | string | — | Name of Principal Investigator |
| PROJECT_CO_I | string | — | Name(s) of the Co-investigator(s) |
| PROJECT_CONTACT | string | — | Contact details for project |
| OBSERVATION_ID | string | — | Unique identifier for the observation |
| OBSERVATION_START_MJD | string | — | Observation start date (MJD) |
| OBSERVATION_START_TAI | string | — | Observation start date (TAI) |
| OBSERVATION_START_UTC | string | — | Observation start date (UTC) |
| OBSERVATION_END_MJD | string | — | Observation end date (MJD) |
| OBSERVATION_END_TAI | string | — | Observation end date (TAI) |
| OBSERVATION_END_UTC | string | — | Observation end date (UTC) |
| OBSERVATION_NOF_STATIONS | integer | — | nof. stations used during the observation |
| OBSERVATION_STATIONS_LIST | array<string,1> | — | List of stations used during the observation |
| OBSERVATION_FREQUENCY_MAX | double | — | Observation maximum frequency |
| OBSERVATION_FREQUENCY_MIN | double | — | Observation minimum frequency |
| OBSERVATION_FREQUENCY_CENTER | double | — | Observation center frequency |
| OBSERVATION_FREQUENCY_UNIT | string | 'MHz' | Frequency units of this observation |
| OBSERVATION_NOF_BITS_PER_SAMPLE | int | — | Number of bits per sample in the incoming data stream from the stations to CEP/BlueGene. |
| CLOCK_FREQUENCY | double | — | Clock frequency, in units of CLOCK_FREQUENCY_UNIT; valid values for LOFAR are 160.0 MHz and 200.0 MHz. |
| CLOCK_FREQUENCY_UNIT | string | 'MHz' | Clock frequency unit |
| ANTENNA_SET | string | — | Antenna set specification of observation |
| FILTER_SELECTION | string | — | Filter selection (see description) |
| TARGET | string | — | Single or list of observation targets/-sources |
| SYSTEM_VERSION | string | — | Processing system name/version |
| PIPELINE_NAME | string | — | Pipeline processing name |
| PIPELINE_VERSION | string | — | Pipeline processing version |
| NOTES | string | — | Notes or comments |

Table 2: Common LOFAR Attributes (CLA)

| 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) intensity as a function of frequency, or (2) voltage vs time. |

Table 3: Overview of standard LOFAR data products and the corresponding file type attribute value.

– `TELESCOPE` - **name of the telecope** with which the observation was carried out – i.e. `LOFAR`.

– `OBSERVER` - holds the **name(s) of the observer(s)**.

– If the observation is carried out within the context of a specific **project**, then its ID will be stored in `PROJECT_ID` and title within `PROJECT_TITLE`. Additional attributes provide further detailed information, such as the name of the project's principal investigator (`PROJECT_PI`), the name(s) of the co-investigator(s) (`PROJECT_CO_I`) as well as means to contact the project (`PROJECT_CONTACT`). If no specific project is defined, the variables simply should be set to `'LOFAR'`.

– `OBSERVATION_ID` – is the **unique identifier** for the LOFAR observation.

– The observation's start time is listed in the following formats:
  - Modified Julian Day (`OBSERVATION_START_MJD`),
  - International Atomic Time (`OBSERVATION_START_TAI`) and
  - Coordinated Universal Time (`OBSERVATION_START_UTC`).

– The observation's end time is listed in the following formats:
  - Modified Julian Day (`OBSERVATION_END_MJD`),
  - International Atomic Time (`OBSERVATION_END_TAI`) and
  - Coordinated Universal Time (`OBSERVATION_END_UTC`).

– `OBSERVATION_NOF_STATIONS` – Number of stations used for this observation

– `OBSERVATION_STATIONS_LIST` – A list of stations used for this observation

– `OBSERVATION_FREQUENCY_MAX` – Upper frequency limit of observation data

– `OBSERVATION_FREQUENCY_MIN` – Lower frequency limit of observation data

– `OBSERVATION_FREQUENCY_CENTER` – Center frequency of the covered frequency range, given as the geometric mean of maximum and minimum frequency:

$$\begin{aligned} \nu_{\rm center} &= (\nu_{\rm min} + \nu_{\rm max})/2 \\ &= (\texttt{OBSERVATION\_FREQUENCY\_MIN} + \texttt{OBSERVATION\_FREQUENCY\_MAX})/2 \end{aligned}$$

Given the possibilities of rather non-regular coverage in frequency space, $\nu_{\rm center}$ is formost intended as orientation during the initial inspection of the data sets' properties; for precise information on the sampling in frequency space, one is refered to the Spectral coordinate as part of the Coordinates group.

- OBSERVATION_FREQUENCY_UNIT – When TELESCOPE is 'LOFAR', all observation frequency units will be 'MHz'.
- CLOCK_FREQUENCY – The clocking frequency used for the observation. For LOFAR, this will be one of '160' or '200'.
- CLOCK_FREQUENCY_UNIT – For LOFAR, this will be 'MHz'
- ANTENNA_SET – The **antenna set** configuration used during the observation; see Table 4 below for a list of recognized values.

| Antenna set | Description |
|---|---|
| 'LBA_INNER' | 48 antennas of the INNER LBA configuration (see figure 2) |
| 'LBA_OUTER' | 48 antennas of the OUTER LBA configuration (see figure 2) |
| '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, 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. For Core stations, this will result in a "weird" beamshape. |

Table 4: Overview of antenna set configurations.

- FILTER_SELECTION – The **filter selection** (frequency bandwidth) used during the observation. The metadata need to reflect the frequency band in which the data have been recorded; see Table 5 below for a list of recognized values.

| 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.

- TARGET - User-supplied target name holds a single source name or a list of the observed sources/targets. This field can also state that the observation was 'All-sky' or reference a grid number/identifier as part of an all-sky survey.
- SYSTEM_VERSION lists the name and (if available) version of the processing system used for carying out the observation and creating the data.
- PIPELINE_NAME and PIPELINE_VERSION list name and version of the pipeline by which the data have been processed to the recorded state.

– The NOTES attributes acts as generic area for notes and comments.

| **General LOFAR Group** | **Value** | **Description** |
|---|---|---|
| Root | 'Root' | Top-level LOFAR group type |
| System Log | 'SysLog' | System log files, parsets |
| **Image** | 'Image' | Image group |
| **Image Group Subgroups** | **Value** | **Description** |
| Data group (DATA) | 'Data' | This is a Sky Image dataset. |
| Source table (SOURCE_TABLE) | 'SourceTable' | This is a Source table. |
| Processing History group | 'ProcessHistory' | This is a Processing History group |
| **Coordinate Groups** | **Value** | **Description** |
| Coordinates Group | 'Coordinates' | This is a Coordinates group |
| Direction coord group | 'DirectionCoord' | This is a direction coordinate group |
| Spectral coord group | 'SpectralCoord' | This is a frequency coordinate group |
| Polarisation coord group | 'PolarisationCoord' | This is a polarisation coordinate group |

Table 6: LOFAR Sky Image Group Types

### 4.1.2. Additional Sky Image Root Attributes

As explained at the begin of Sec. 4.1 above, the root group of a **Radio Sky Image Cube** 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 Sky Image data. With the Common LOFAR Attributes already listed in Sec. 4.1.1 above, this section will focus on the second subset of root group attributes, as they are specific to a **Radio Sky Image Cube**.

This root group header will comprise general information about the observation itself, sparing relevant data details for the headers of the lower order sub-groups. Table 7 presents additional root group attributes for a LOFAR **Radio Sky Image Cube** that do not appear in the Common LOFAR Attributes table.[1]

| Field/Keyword | Type | Description |
|---|---|---|
| INPUT_DATA | string | Descriptor to track the input data source, from which the image was created; this can be a visibility dataset, but as well streaming data (where no intermediate data product is being produced). |
| NOF_IMAGES | int | N Image groups in this file |
| TARGET_RA | float | RA of TARGET (at LOFAR core) |
| TARGET_DEC | float | Dec of TARGET (at LOFAR core) |

Table 7: Additional root group attributes specific to a Radio Sky Image Cube

## 4.2. The System Logs Group (SYS_LOG)

[ **Comment:**
This is where the SYS-LOG Group should be described. ]

[1] * Indicates attributes that may *migrate from the root group* and be broadcast to individual Image groups. Recent observations have indicated that different sub-bands potentially can have different integration times.

## 4.3. The Image Group (IMAGE_{NNN})

The `Image` group will be an HDF5 group serving as a container for the four sub groups described below. An `Image` group is designed to be as complete and self-contained an image cube as possible, and will contain relevant data and metadata for a particular processed sub-band of a LOFAR observation. However, any breakout protocol will be required to inherit some or all root group attributes in order to function as a stand-alone image. The adopted form allows for relatively simple extraction and conversion in a FITS-compatible form.

An `Image` group will comprise four sub groups. These four groups, two hierarchical levels below the "root group" in a LOFAR Sky Image (Fig. 1), will be

- A `COORDINATES` group that will contain one or more subgroups of kinds `LinearCoord, TabularCoord, SpectralCoord, DirectionCoord, StokesCoord`, that will describe various axes of the associated dataset.
- A `IMAGE_DATA` dataset that will contain a dataset array.
- A `SOURCE_TABLE` that will be a tabular representation of a Local Sky Model, as well as clean component used for the deconvolution of the image.
- A `PROCESS_HISTORY` group, which will be a meta-data container holding various processing products such as log files, parameter sets, RFI mitigation tables, etc.

The diagram below illustrates the form of an Image group in a LOFAR Sky Image file. The table of Image group attributes is notably sparse here and the reader must bear in mind that the Coordinates groups will contain most of the rest of the relevent Image group metadata (*see 4.4, "Coordinates group"*)

```
.
|-- GROUPTYPE              ...  Attr.          string
|-- COORDINATES            ...  Group
|   |-- COORDINATE_0       ...  Group
|   |-- COORDINATE_1       ...  Group
|   `-- COORDINATE_2       ...  Group
|-- IMAGE_DATA             ...  Dataset
|-- SOURCE_TABLE           ...  Dataset
`-- PROCESS_HISTORY        ...  Group
```

| Field/Keyword | H5Type | Type | Value | Description |
|---|---|---|---|---|
| GROUPTYPE | Attribute | string | 'Image' | LOFAR group type |
| COORDINATES | Group | — | — | Coordinates group |
| IMAGE_DATA | Dataset | — | — | Image dataset |
| SOURCE_TABLE | Dataset | — | — | Source list table/dataset |
| PROCESS_HISTORY | Group | — | — | Processing history group |

Table 8: Components attached to the image group.

## 4.4. The Coordinates Group (COORDINATES)

Coordinate information within a LOFAR **Radio Sky Image Cube** will exist in what is called a **Coordinates group**, which will act as a container for a number of Coordinates group objects. The **Coordinates group** (`COORDINATES`) will be a subgroup of an **Image group** container, and may contain one or more subgroups that will describe relevent axes of the coordinates' associated `Data` group using one or a combination of coordinate subgroups, where the enumerated are Direction coordinate, Spectral coordinate and Polarization coordinate.

The attributes, as presented in Table 9, summarize the overall characteristics of the set of coordinates collected within this group:

| Field/Keyword | Type | Description |
|---|---|---|
| GROUPTYPE | string | Group type descriptor, Coordinates |
| REF_LOCATION_VALUE | array<double,1> | Numerical value(s) of the reference location |
| REF_LOCATION_UNIT | array<string,1> | Physical unit(s) for the reference location |
| REF_LOCATION_FRAME | string | Identifier for the reference system of the location; see Tab. 10 for a list of recognized values. |
| REF_TIME_VALUE | double | Numerical value of the reference time |
| REF_TIME_UNIT | string | Physical unit of the reference time |
| REF_TIME_FRAME | string | Identifier for the reference time system used |
| NOF_COORDINATES | int | N of coordinate objects |
| NOF_AXES | int | N of coordinate axes |
| COORDINATE_TYPES | array<string,1> | embedded coordinate object types |
| COORDINATE_{N} | Group | coordinate object container |

Table 9: Components of a Coordinates group.

```
.
`- COORDINATE_{N}             Group
   |- GROUPTYPE               Attr.        string
   |- COORDINATE_TYPE         Attr.        string
   |- STORAGE_TYPE            Attr.        string
   |- NOF_AXES                Attr.        int
   |- AXIS_NAMES              Attr.        array<string,1>
   |- AXIS_UNITS              Attr.        array<string,1>
   |- REFERENCE_VALUE         Attr.        array<double,1>
   |- REFERENCE_PIXEL         Attr.        array<double,1>
   |- INCREMENT               Attr.        array<double,1>
   |- PC                      Attr.        array<double,1>
   |- EQUINOX                 Attr.        string
   |- RADEC_SYS               Attr.        string
   |- PROJECTION              Attr.        string
   |- PROJECTION_PARAM        Attr.        array<double,1>
   |- LONGPOLE                Attr.        double
   `- LATPOLE                 Attr.        double
```

Listing 1: Structure of the direction coordinate group.

- GROUPTYPE – Identifier for the type of group, “Coordinates”.
- NOF_COORDINATES – The number of coordinate objects/groups contained within the coordinates group.
- NOF_AXES – The number of coordinate axes associated with the coordinate objects. Keep in mind, that a coordinate can have multiple (coupled) axes: e.g. a direction coordinate is composed of two axes.

The layout of the embedded sub-groups will depend on the type of coordinate, of which there are several types.

#### 4.4.1. Direction coordinate

The Direction Coordinate consists of a set of two coupled coordinate axes, describing a direction in space; it therefore includes information such as the equinox of the observation, the system of equatorial coordinates on the sphere of the sky, as well as parameters for the spherical map projection.

– GROUPTYPE is the group type descriptor with the fixed value ‘DirectionCoord’.

– COORDINATE_TYPE is the is the descriptor for the coordinate type, of value ‘Direction’.

– STORAGE_TYPE is the descriptor for the underlying storage type for this coordinate, of value ‘Direction’.

| Reference Position | Description | Comments |
|---|---|---|
| GEOCENTER | Center of the Earth. | |
| BARYCENTER | Center of the solar system barycenter. | |
| HELIOCENTER | Center of the Sun. | |
| TOPOCENTER | "Local"; in most cases this will mean: the location of the telescope. | |
| LSRK | Kinematic Local Standard of Rest: 20 km s$^{-1}$ in the direction of GALACTIC_II $(56, +23)$. | Only to be used for redshifts and Doppler velocities, and spectral coordinate. |
| LSRD | Dynamic Local Standard of Rest: 16.6 km s$^{-1}$ in the direction of GALACTIC_II $(53, +25)$. | |
| GALACTIC | Center of the Galaxy: 220 km s$^{-1}$ in the direction of GALACTIC_II $(90, 0)$ w.r.t. LSRD. | |
| LOCAL_GROUP | Center of the Local Group: 300 km s$^{-1}$ in the direction of GALACTIC_II $(90, 0)$ w.r.t. BARYCENTER. | |
| RELOCATABLE | Relocatable center; for simulations. | Only to be used for spatial coordinates. |

Table 10: Recognized values for the reference frame to specify a location; values and descriptions have been adopted from the "Space-Time Coordinate Metadata for the Virtual Observatory" [?], as produced by the IVOA Data Model Working Group.

- NOF_AXES is the number of coordinate axes; keep in mind that a coordinate can consist of multiple axes. For the the DirectionCoordinate we have NOF_AXES=2.
- AXIS_NAMES are the world axis names connected with the coordinate axes,most commonly

  AXIS_NAME=['Longitude','Latitude']
- AXIS_UNITS are the physical units along each coordinate axis (corresponding to the FITS keyword CUNITn, see [?]). Restrictions on the nature and range of units, if any, will be determined by agreements applying to the specific axis. If they are not so limited, units should conform to the IAU Style Manual [?].
- REFERENCE_VALUE is the coordinate value at the reference point (corresponding to the FITS keyword CRVAL$n$, see [?]).
- REFERENCE_PIXEL is the array location of the reference point in pixels (corresponding to the FITS keyword CRPIX$n$, see [?]).
- INCREMENT is the coordinate increment at the reference point (corresponding to the FITS keyword CDELT$n$, see [?]).
- PC is a non-singular square matrix, for the transformation from intermediate pixel coordinates to intermediate world coordinates. The individual matrix elements are stored as a linear array, ordered as follows:

$$\mathbf{M}_{[N,N]} = \begin{pmatrix} M_{00} & M_{01} & ... & M_{0N} \\ M_{10} & M_{11} & ... & M_{1N} \\ \vdots & & & \vdots \\ M_{N0} & & & M_{NN} \end{pmatrix} \rightarrow [M_{00}, M_{01}, ..., M_{10}, M_{11}, ..., M_{N0}, ..., M_{NN}]$$

| FIELD/KEYWORD | TYPE | VALUE | DESCRIPTION |
|---|---|---|---|
| GROUPTYPE | string | 'DirectionCoord' | Group type descriptor |
| COORDINATE_TYPE | string | 'Direction' | Coordinate Type descriptor |
| STORAGE_TYPE | string | 'Direction' | Descriptor for the underlying storage type for this coordinate |
| NOF_AXES | int | $N \equiv 2$ | Number of coordinate axes |
| AXIS_NAMES | array<string,1> | $[name_0, name_1]$ | World axis names |
| AXIS_UNITS | array<string,1> | $[unit_0, unit_1]$ | Physical units along each coordinate axis. |
| REFERENCE_VALUE | array<double,1> | $[val_0, val_1]$ | Coordinate value at the reference point |
| REFERENCE_PIXEL | array<double,1> | $[pix_0, pix_1]$ | Array location of the reference point in pixels. |
| INCREMENT | array<double,1> | $[incr_0, incr_1]$ | Coordinate increment at reference point. |
| PC | array<double,1> | $[pc_{00}, pc_{01}, pc_{10}, pc_{11}]$ | Non-singular square matrix, for the transformation from intermediate pixel coordinates to intermediate world coordinates. |
| EQUINOX | string | | Equinox of the observation |
| RADEC_SYS | string | | System of equatorial coordinates |
| PROJECTION | string | | Spherical map projection |
| PROJECTION_PARAM | array<double,1> | | Spherical projection parameters |
| LONPOLE | double | | Native longitude of the celestial pole, $\phi_p$ |
| LATPOLE | double | | Native latitude of the celestial pole, $\theta_p$. |

Table 11: Attributes/keywords attached to a group describing a direction coordinate.

- EQUINOX applies to ecliptic as well as to equatorial coordinates (e.g. J2000 or B1950) of the source position.
- RADEC_SYS Several systems of equatorial coordinates (right ascension and declination) are in common use. Apart from the International Celestial Reference System (ICRS, IAU, 1984), the axes of which are by definition fixed with respect to the celestial sphere, each system is parameterized by time. In particular, mean equatorial coordinates are defined in terms of the epoch (i.e. instant of time) of the mean equator and equinox (i.e. pole and origin of right ascension). The same applies for ecliptic coordinate systems. The keyword RADEC_SYS is used to specify the particular system; recognized values are given in Tab. 12 below.

| RADEC_SYS | DESCRIPTION |
|---|---|
| ICRS | International Celestial Reference System |
| FK5 | mean place, new (IAU 1984) system |
| FK4 | mean place, old (Bessell-Newcomb) system |
| FK4-NO-E | meanplace, old system but without e-terms |
| GAPPT | Geocentric Apparent Place, IAU 1984 system |

Table 12: Allowed values of RADEC_SYS

- PROJECTION holds the reference code for the spherical map projection, e.g. AIT, SIN, STG, etc. As some of these projections require (or at least allow) additional parameters, the PROJECTION_PARAM keyword is used to store these additional parameters. Recognized values are given in Table 13 below.
- LONPOLE is the native longitude of the celestial pole, $\phi_p$.

| | Projection | $\phi_0$ | $\theta_0$ | Projection parameters |
|---|---|---|---|---|
| AZP | Zenithal perspective | 0° | 90° | $[\mu, \gamma]$ |
| SZP | Slant zenithal perspective | 0° | 90° | $[\mu, \phi_c, \theta_c]$ |
| TAN | Gnomonic | 0° | 90° | |
| STG | Stereographic | 0° | 90° | |
| SIN | Slant orthographic | 0° | 90° | $[\xi, \eta]$ |
| ARC | Zenithal equidistant | 0° | 90° | |
| ZPN | Zenithal polynomial | 0° | 90° | $[P_0, P_1, .., P_m]$ for $m = 0, \dots 29$ |
| ZEA | Zenithal equal-area | 0° | 90° | |
| AIR | Airy | 0° | 90° | $[\theta_b]$ |
| CYP | Cylindrical perspective | 0° | 0° | $[\mu, \lambda]$ |
| CEA | Cylindrical equal area | 0° | 0° | $[\lambda]$ |
| CAR | Plate carrée | 0° | 0° | |
| MER | Mercator | 0° | 0° | |
| SFL | Sanson-Flamsteed | 0° | 0° | |
| PAR | Parabolic | 0° | 0° | |
| MOL | Mollweide | 0° | 0° | |
| AIT | Hammer-Aitoff | 0° | 0° | |
| COP | Conic perspective | 0° | $\theta_a$ | $[\theta_a, \eta]$ |
| COE | Conic equal-area | 0° | $\theta_a$ | $[\theta_a, \eta]$ |
| COD | Conic equidistant | 0° | $\theta_a$ | $[\theta_a, \eta]$ |
| COO | Conic orthomorphic | 0° | $\theta_a$ | $[\theta_a, \eta]$ |
| BON | Bonne's equal area | 0° | 0° | $[\theta_1]$ |
| PCO | Polyconic | 0° | 0° | |
| TSC | Tangential Spherical Cube | 0° | 0° | |
| CSC | COBE Quadrilateralized Spherical Cube | 0° | 0° | |
| QSC | Quadrilateralized Spherical Cube | 0° | 0° | |

Table 13: Summary of projection codes, full name, default values of $\phi_0$ and $\theta_0$, and required parameters. Values and descriptions have been adopted from [?].

– LATPOLE is the native latitude of the celestial pole, $\theta_p$.

#### 4.4.2. Spectral coordinate

Spectral coordinates are commonly given in units of frequency, wavelength, velocity, and other parameters proportional to these three [?]. The coordinate types discussed here are then frequency, wavelength, and apparent radial velocity denoted by the symbols $\nu$, $\lambda$, and $v$. There are also three conventional velocities frequently used in astronomy. These are the so-called radio velocity, optical velocity, and redshift, denoted here by $V$, $Z$, and $z$ and given by

$$V = c\,\frac{\nu_0 - \nu}{\nu_0}\,, \qquad Z = c\,\frac{\lambda - \lambda_0}{\lambda_0} \quad \text{and} \qquad z = Z/c\,.$$

The velocities are defined so that an object receding from the observer has a positive velocity. Table 14 below lists the various spectral quantities and their respective encoding as an attribute; the symbols $\lambda_0$ and $\nu_0$ are the rest wavelength and frequency, respectively, of the spectral line used to associate velocity with observed wavelength and frequency.

As it turns out, providing a set of parameters to properly describe a spectral coordinate is not straightforward: given the arrangement of frequency channels or bands the values along the coordinate axis might be linear, but does not necessarily have to be. Therefore in principle a spectral coordinate can be considered

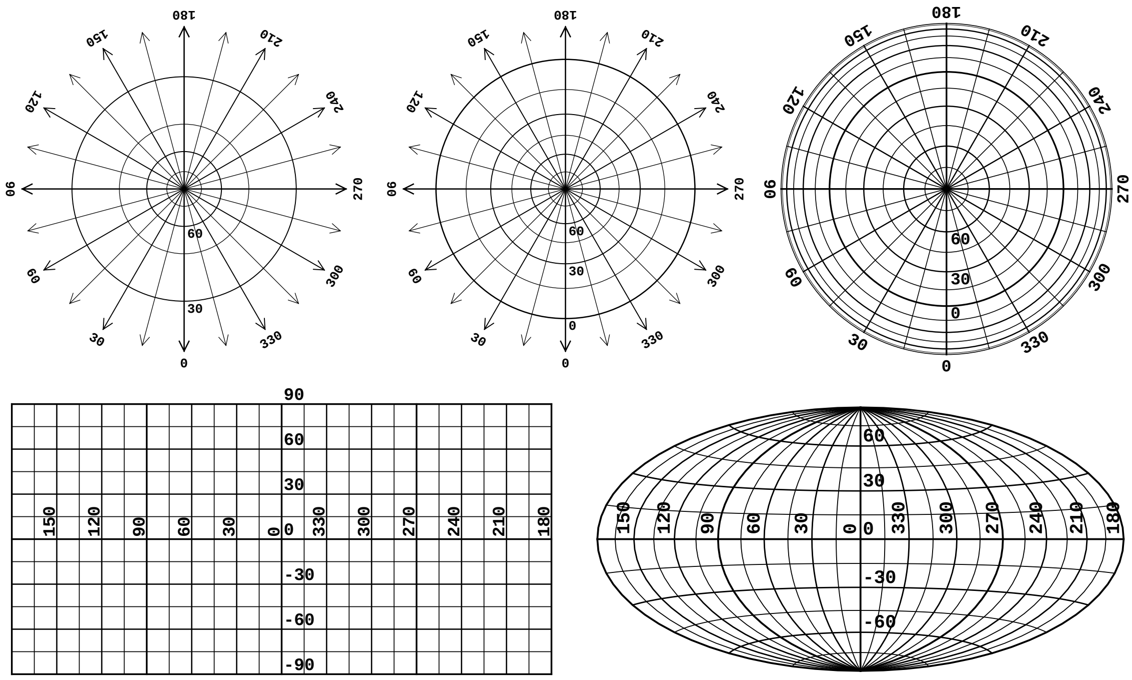

Figure 2: A selection of spherical map projections. Top row, from left to right: `TAN` (Gnomonic), `STG` (Stereographic), `ZEA` (Zenithal equal-area). Bottom row, from left to right: `CAR` (Plate carrée), `AIT` (Hammer-Aitoff).

a derivative of either a linear or a tabular coordinate, with a number of specific attributes added, as they will be required for the transformation between different spectral quantities.

– `GROUPTYPE` is the group type descriptor with the fixed value `‘SpectralCoord’`.

– `COORDINATE_TYPE` is the coordinate type descriptor with the fixed value ‘SPECTRAL’.

– `NOF_AXES` is the number of coordinate axes.

– `AXIS_NAMES` are the world axis names associated with the spectral coordinates, e.g.

  `AXIS_NAMES=[‘Frequency’]`
  `AXIS_NAMES=[‘WavelengthVacuum’]`

  Allowed and supported values are listed in Tab. 14 above.

– `AXIS_UNITS` are the physical units along each coordinate axis (corresponding to the FITS keyword `CUNITi`, see [?]). Restrictions on the nature and range of units, if any, will be determined by agreements applying to the specific axis. If they are not so limited, units should conform to the IAU Style Manual [?].

– `STORAGE_TYPE` indicates the underlying storage mechanism: if `STORAGE_TYPE=‘Linear’` the coordinate axis is expected to be linear and represented by the attributes defined for a Linear Coordinate:

```
.
`- COORDINATE_{N}            Group
   |- GROUPTYPE              Attr.        string
   |- COORDINATE_TYPE        Attr.        string
   |- STORAGE_TYPE           Attr.        string
   |- REFERENCE_FRAME        Attr.        string
   |- REST_FREQUENCY         Attr.        double
   |- REST_FREQUENCY_UNIT    Attr.        string
```

| Attribute | FITS Code | Name | Symbol | Associate variable | Default units |
|---|---|---|---|---|---|
| `Frequency` | `FREQ` | Frequency | $\nu$ | $\nu$ | Hz |
| `Energy` | `ENER` | Energy | $E$ | $\nu$ | J |
| `Wavenumber` | `WAVN` | Wavenumber | $\kappa$ | $\nu$ | $\mathrm{m}^{-1}$ |
| `VelocityRadio` | `VRAD` | Radio velocity | $V$ | $\nu$ | $\mathrm{m\ s}^{-1}$ |
| `VelocityOptical` | `VOPT` | Optical velocity | $Z$ | $\lambda$ | $\mathrm{m\ s}^{-1}$ |
| `VelocityAppRadial` | `VELO` | Apparent radial velocity | $v$ | $v$ | $\mathrm{m\ s}^{-1}$ |
| `Redshift` | `ZOPT` | Redshift | $z$ | $\lambda$ | – |
| `WavelengthVacuum` | `FREQ` | Vacuum wavelength | $\lambda$ | $\lambda$ | m |
| `WavelengthAir` | `AWAV` | Air wavelength | $\lambda_a$ | $\lambda_a$ | m |
| `BetaFactor` | `BETA` | Beta factor $v/c$ | $\beta$ | $v$ | – |

Table 14: Attributes values corresponding to the spectral coordinate codes, as defined in [?]. The IAU-standard prefixes for scaling the unit are described in [?] and should be used with al coordinate types, except that the dimensionless ones are not scaled.

| Name | Symbol | Transformation equation(s) |
|---|---|---|
| Frequency | $\nu$ | $\nu = c/\lambda = E/h$ |
| Vacuum wavelength | $\lambda$ | $\lambda = c/\nu = \lambda_0 \dfrac{c+v}{\sqrt{c^2-v^2}}$ |
| Apparent radial velocity | $v$ | $v = c\,\dfrac{\nu_0^2-\nu^2}{\nu_0^2+\nu^2} = c\,\dfrac{\lambda^2-\lambda_0^2}{\lambda^2+\lambda_0^2}$ |
| Energy | $E$ | $E = h\nu$ |
| Redshift | $z$ | $z = \dfrac{\lambda-\lambda_0}{\lambda}$ |

Table 15: Spectral transformation equations; for the full set of equations – including first order derivatives – see [?].

```
|- REST_WAVELENGTH        Attr.       double
|- REST_WAVELENGTH_UNIT   Attr.       string
|- NOF_AXES               Attr.       int
|- AXIS_NAMES             Attr.       array<string,1>
|- AXIS_UNITS             Attr.       array<string,1>
|- REFERENCE_VALUE        Attr.       array<double,1>
|- REFERENCE_PIXEL        Attr.       array<double,1>
|- INCREMENT              Attr.       array<double,1>
`- PC                     Attr.       array<double,1>
```

In this:

- `REFERENCE_VALUE` is the coordinate value at the reference point (corresponding to the FITS keyword `CRVAL`$n$, see [?]).
- `REFERENCE_PIXEL` is the array location of the reference point in pixels (corresponding to the FITS keyword `CRPIX`$n$, see [?]).
- `INCREMENT` is the coordinate increment at the reference point (corresponding to the FITS keyword `CDELT`$n$, see [?]).
- `PC` is a non-singular square matrix, for the transformation from intermediate pixel coordinates to intermediate world coordinates. The individual matrix elements are stored as a linear array,

| FIELD/KEYWORD | TYPE | VALUE | DESCRIPTION |
|---|---|---|---|
| GROUPTYPE | string | 'SpectralCoord' | Group type descriptor |
| COORDINATE_TYPE | string | 'Spectral' | Coordinate Type descriptor |
| STORAGE_TYPE | array<string,1> | 'Linear' \| 'Tabular' | Descriptor for the underlying storage type for this coordinate |
| REFERENCE_FRAME | string | | Reference position w.r.t. which the spectral coordinate axis are defined; see Tab. 10 for a list of recognized values. This can be a diferent frame as used for e.g. the direction coordinate or as noted in the coordinates group. |
| REST_FREQUENCY | double | | Rest frequency, $\nu_0$ |
| REST_FREQUENCY_UNIT | string | 'Hz' | Physical units within which the rest frequency is given |
| REST_WAVELENGTH | double | | Rest wavelength, $\lambda_0$ |
| REST_WAVELENGTH_UNIT | string | 'm' | Physical units within which the rest wavelength is given |
| NOF_AXES | int | $N \equiv 1$ | Number of coordinate axes |
| AXIS_NAMES | array<string,1> | $[name_0]$ | World axis names |
| AXIS_UNITS | array<string,1> | $[unit_0]$ | Physical units along each coordinate axis. |
| REFERENCE_VALUE | array<double,1> | $[val_0]$ | Coordinate value at the reference point |
| REFERENCE_PIXEL | array<double,1> | $[pix_0]$ | Array location of the reference point in pixels. |
| INCREMENT | array<double,1> | $[incr_0]$ | Coordinate increment at reference point. |
| PC | array<double,1> | $[p_{00}] \equiv 1$ | Non-singular square matrix, for the transformation from intermediate pixel coordinates to intermediate world coordinates. |
| AXIS_LENGTH | int | $N_{\text{Pixels}}$ | Length of the axis, i.e. the number of elements stored in the AXIS_VALUES_PIXEL and AXIS_VALUES_WORLD arrays. |
| AXIS_VALUES_PIXEL | array<double,1> | $[p_0, .., p_{N_{\text{Pixels}}}]$ | Tabulated values along the pixel axis. |
| AXIS_VALUES_WORLD | array<double,1> | $[w_0, .., w_{N_{\text{Pixels}}}]$ | Tabulated values along the world axis. |

Table 16: Keywords decribing a Spectral Coordinate; attributes within the first segment of the table will be present independent of the specific storage method.

ordered as follows:

$$\mathbf{M}_{[N,N]} = \begin{pmatrix} M_{00} & M_{01} & ... & M_{0N} \\ M_{10} & M_{11} & ... & M_{1N} \\ \vdots & & & \vdots \\ M_{N0} & & & M_{NN} \end{pmatrix} \rightarrow [M_{00}, M_{01}, ..., M_{10}, M_{11}, ..., M_{N0}, ..., M_{NN}]$$

If set STORAGE_TYPE='Tabular', the values along the coordinate axis are expected to be tabulated, thereby represented by the attributes defined for a Tabular Coordinate:

```
.
`- COORDINATE_{N}            Group
   |- GROUPTYPE              Attr.       string
   |- COORDINATE_TYPE        Attr.       string
   |- STORAGE_TYPE           Attr.       string
   |- REFERENCE_FRAME        Attr.       string
   |- REST_FREQUENCY         Attr.       double
   |- REST_FREQUENCY_UNIT    Attr.       string
   |- REST_WAVELENGTH        Attr.       double
   |- REST_WAVELENGTH_UNIT   Attr.       string
   |- NOF_AXES               Attr.       int
   |- AXIS_NAMES             Attr.       array<string,1>
   |- AXIS_UNITS             Attr.       array<string,1>
   |- AXIS_LENGTH            Attr.       int
   |- AXIS_VALUES_PIXEL      Attr.       array<double,1>
   `- AXIS_VALUES_WORLD      Attr.       array<double,1>
```

In this:

- AXIS_LENGTH is the length of the tabulated axis, i.e. the number of elements stored in the AXIS_VALUES_PIXEL and AXIS_VALUES_WORLD arrays.
- AXIS_VALUES_PIXEL are the tabulated values along the pixel axis.
- AXIS_VALUES_WORLD are the tabulated values along the world axis.

### 4.4.3. Polarization coordinate

| Field/Keyword | Type | Value | Description |
|---|---|---|---|
| GROUPTYPE | string | 'PolarizationCoord' | Group type descriptor |
| COORDINATE_TYPE | string | 'Polarization' | Coordinate Type descriptor |
| STORAGE_TYPE | array<string,1> | 'Tabular' | Descriptor for the underlying storage type for this coordinate |
| NOF_AXES | int | $N \equiv 1$ | Number of coordinate axes |
| AXIS_NAMES | array<string,1> | $[name_0] \equiv$ 'Polarization' | World axis names |
| AXIS_UNITS | array<string,1> | $[unit_0]$ | Physical units along each coordinate axis. |
| AXIS_LENGTH | int | $N_{\text{Length}}$ | Length of the axis, i.e. the number of elements stored in the AXIS_VALUES_PIXEL and AXIS_VALUES_WORLD arrays. |
| AXIS_VALUES_PIXEL | array<int,1> | $[p_0, .., p_{N_{\text{Length}}}]$ | Tabulated values along the pixel axis. |
| AXIS_VALUES_WORLD | array<string,1> | $[w_0, .., w_{N_{\text{Length}}}]$ | Tabulated values along the world axis, listing the stored Polarization parameters. |

Table 17: Keywords decribing a Polarization Coordinate.

– GROUPTYPE is the group type descriptor with the fixed value 'PolarizationCoord'.

– COORDINATE_TYPE is the is the descriptor for the coordinate type, of value 'Polarization'.

– STORAGE_TYPE is the descriptor for the underlying storage type for this coordinate, of value 'Tabular'.

– NOF_AXES is the number of coordinate axes represented by this coordinate; as the Polarization coordinate consists of a single tabulared axis, we have NOF_AXES = 1.

– AXIS_NAMES are the world axis names connected with the coordinate axes; for a Polarization coordinate AXIS_NAMES = 'Polarization'.

– AXIS_UNITS are the physical units along each coordinate axis (corresponding to the FITS keyword CUNITi, see [?]). Restrictions on the nature and range of units, if any, will be determined by agreements applying to the specific axis. If they are not so limited, units should conform to the IAU Style Manual [?].
The units of the Stokes parameters I, Q, U and V, of total polarization (linear, elliptical or circular) and of separate circular polarizations (L, R) are some form of flux density.

– AXIS_VALUES_PIXEL holds the tabulated values along the pixel axis

– AXIS_VALUES_WORLD holds the tabulated values along the world axis of the Polarization coordinate, i.e. the names of the Polarization components. Commonly used values are:

| AXIS_VALUES_WORLD | DESCRIPTION |
|---|---|
| ['I'] | Total flux density only data. |
| ['I','Q','U','V'] | Full set of standard Stokes parameters. |
| ['X','Y'] | Raw time-series TBB data, originating directly from the individual dipoles. |
| ['XX','YY','XY','YX'] | Cross-correlation products from a pair of $X$-linear and $Y$-linear receiver feeds. |
| ['R','L','X','Y'] | $X/Y$ linear components, as well as $R/L$ circular components. |

For a full list of recognized values and their description see Tab. 18 below.

## 4.5. The Image Dataset (DATA)

The **Image Dataset** is the object within the **Radio Sky Image Cube** structure responsible for the actual storage of the image pixel data. As such the dataset is a multi-dimensional array, storing intensity values as function of direction on the sky (Lon, Lat), frequency ($\nu$) and polarization component ($p$):

$$I = I(p, \nu, \mathrm{Lat}, \mathrm{Lon})$$

Attached to the data array are a small number of attributes – see Tab. 19 below – describing basic properties such as array dimensions.

## 4.6. The Source Table (SOURCE_TABLE)

The Source table in a LOFAR **Radio Sky Image Cube** will collect sources and their associated parameters, as extracted from either Local Sky Model (LSM) or the image data itself. The Source table header will specify the fields (columns) of the table, the number of sources in the table (rows). See Table 20 for the specification of Source group attributes for a LOFAR **Radio Sky Image Cube**.

- GROUPTYPE – Group type of this structure, SourceTable.

- NOF_TABLE_ROWS – Number of sources listed/collected in the source table; with one source per row this corresponds to the number or table rows.

| Term | Symbol | Description |
|---|---|---|
| Stokes Parameters | I | Standard Stokes total intensity, i.e. total Poynting vector or flux density of the wave. |
| | Q | Standard Stokes linear; degree of polarization, i.e. the difference in intensities between horizontal and vertical linearly polarized components. |
| | U | Standard Stokes linear; plane of polarization, i.e. the difference in intensities between linearly polarized components oriented at $\pm\pi/4$ w.r.t. the components of $Q$ |
| | V | Standard Stokes circular; ellipticity, i.e. the differences in intensities between right and left circular polarized components. |
| Circular feeds | R | Right circular |
| | L | Left circular |
| | RR | Right-right circular |
| | LL | Left-left circular |
| | RL | Right-left circular |
| | LR | Left-right circular |
| Linear feeds | X | X linear |
| | Y | Y linear |
| | XX | X parallel linear |
| | YY | Y parallel linear |
| | XY | XY cross linear |
| | YX | YX cross linear |

Table 18: Recognized values for the Polarization component parameter.

| Field/Keyword | Type | Value | Description |
|---|---|---|---|
| `GROUPTYPE` | `string` | `'Data'` | Group type descriptor |
| `WCSINFO` | `string` | `'/Coordinates'` | Path to the coordinates group describing the transformation for array pixel axes to world coordinates. |
| `DATASET_NOF_AXES` | `int` | — | Number of array axes of the dataset |
| `DATASET_SHAPE` | `array<int,1>` | — | Shape of the image data array. |

Table 19: Attributes attached to the image dataset array.

- `NOF_TABLE_COLUMNS` – The number of table columns.
- `TABLE_COLUMNS` – Name of the table columns.
- `EQUINOX` – Equinox of the source position, e.g. `J2000` or `B1950`.
- `RADEC_SYS` Several systems of equatorial coordinates (right ascension and declination) are in common use. Apart from the International Celestial Reference System (ICRS, IAU, 1984), the axes of which are by definition fixed with respect to the celestial sphere, each system is parameterized by time. In particular, mean equatorial coordinates are defined in terms of the epoch (i.e. instant of time) of the mean equator and equinox (i.e. pole and origin of right ascension). The same applies for ecliptic coordinate systems. The keyword `RADEC_SYS` is used to specify the particular system; recognized values are given in Tab. 12 above.
- `RADEC_UNITS` – Physical units – degress (`deg`) or radian (`rad`) – within which the source position is recorded. Instead of `hh:mm:ss` we are using a decimal representation of the celestial position.
- `STOKES_COMPONENTS` – Stokes components for which the source flux is given, e.g. `['I','Q','U','V']`

| Field/Keyword | Type | Value | Description |
|---|---|---|---|
| GROUPTYPE | string | 'SourceTable | LOFAR group type |
| NOF_TABLE_ROWS | int | | Number of table rows. |
| NOF_TABLE_COLUMNS | int | | Number of table columns. |
| TABLE_COLUMNS | array<string,1> | | Name of the table columns. |
| EQUINOX | string | 'J2000' | Equinox of the observation |
| RADEC_SYS | string | 'FK5' | System Ra and Dec |
| RADEC_UNITS | string | | Physical units – degress (deg) or radian (rad) – within which the source position is recorded. |
| STOKES_COMPONENTS | array<string,1> | — | Stokes components for which the source flux is given. |

Table 20: Attributes of the Source table; attributes visible at this level – e.g. definition of reference system for the position – will be shared across all entries within the table.

or ['I','RR','LL']. For a more details overview of possible combinations of Stokes parameters see [?].

| Column/Keyword | Type | Value | Description |
|---|---|---|---|
| NUMBER | int | — | Running index of the table entry. |
| NAME | string | — | Name of the source |
| RA | double | — | RA position of the source. |
| DEC | double | — | Dec position of the source. |
| STOKES_COMPONENTS | array<string,1> | — | Stokes components for which the flux components are listed. |
| FLUX_PEAK | array<double,1> | — | Peak flux of the source, as per Stokes component. |
| FLUX_INTEGRATED | array<double,1> | — | Integrated flux of the source, as per Stokes component. |
| MODEL_TYPE | string | — | Parametric model usedfor the description of the source shape; point source ('Point'), Gaussian(s) ('Gaussian'), Shapelets ('Shapelet') |
| MODEL_PARAMETER_NAMES | array<string,1> | — | Parameters required for the description of the source, according to the specified MODEL_TYPE. |
| MODEL_PARAMETER_VALUES | array<double,1> | — | Parameters required for the description of the source, according to the specified MODEL_TYPE. |

Table 21: Columns of the source table.

## 4.7. The Processing History Group (PROCESS_HISTORY)

The data definition for the **Processing History Group** is necessarily loose, and will accommodate a variety of ancillary meta-data related to or produced by the various LOFAR processing pipelines. Products such as DPPP log files, processing parameter sets, RFI mitigation tables, etc., may appear in this group. In fact, and due to the wide-ranging data types and free-form ASCII format the many log files may present, the **Processing History Group** will be a catch-all envelop encapsulating information about all steps of processing should the user need such information.

As with all other **Radio Sky Image Cube** HDF5 groups and subgroups, the **Processing History Group** will be an HDF5 group, as a subgroup of an Image group. The attributes will contain a brief summary of the appended processing files contained therein, with pointers to tables containing the logging data, parameter sets, etc..

| Field/Keyword | H5Type | Type | Value | Description |
|---|---|---|---|---|
| GROUPTYPE | Attr. | string | 'ProcessHist' | LOFAR group type |
| DPPP_LOG | Attr. | bool | | DPPP process log? |
| DPPP_PARSET | Attr. | bool | | DPPP parset file? |
| IMAGER_LOG | Attr. | bool | | Imager log? |
| IMAGER_PARSET | Attr. | bool | | Image parset file? |
| BBS_LOG | Attr. | bool | | BBS process log? |
| BBS_PARSET | Attr. | bool | | BBS parset? |

Table 22: Attributes of a Processing History group.

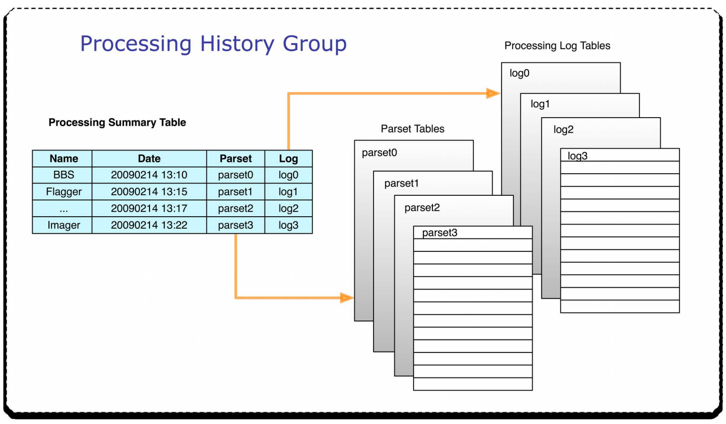


Figure 3: The processing history group (PROCESS_HISTORY), nested tabulation

# 5. Interfaces

— / —

# A. Discussion & open questions

## A.1. Open questions/Issues

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

| Item | Description | Status |
|---|---|---|
| 01 | Check if description of Stokes coordinate is complete | open |
| 02 | Source table (Tab. 4.6): Do we need to track uncertainties as well, i.e. errors on the source position? | open |
| 03 | Source table (Tab. 4.6): How flexible do we need to be with the units and the reference system for the source position? Is there a single setting for all entries, or do we need to allow this to be set for each individual entry? | open |
| 04 | I see references to a "sky image file"n (Sec. 3). What does "file" mean in this context? How does it square with the (oft discussed) capability of HDF5 to agglomerate multiple independent entities into a single data structure? Is this really a file on disk, or is it potentially some collection of files, perhaps even on different hosts? | open |
| 05 | The implicit assumption here is that an image cube contains the data for all frequencies for a given observation (start, end) time. That's not necessarily a bad idea, but it's not obvious why it should be the case. One could imagine an equally valid structure that contained all timesteps for a given (start time, frequency) pair. Or, indeed, a hypercube that contained both time and frequency axes. Indeed, since the individual image groups are basically independent (sharing only a system log group), why should they come from the same observation at all? Would it be useful to, for example, be able to make a sky image cube that contained images from **all** observations of a given source in one coherent data structure? | open |
| 06 | It's likely that there will be a complex structure in the data which makes up an individual subband image. Consider just a 64 channel subband over just a single correlator dump. It's possible that some channels will be flagged out before imaging. What is recorded as the frequency of the resultant image? The central frequency of the unflagged subband? The average frequency of the unflagged channels? Of course, a similar question applies to time: what if some timesteps are flagged in a given subband? Even more complex, it's likely that some channels will only be flagged at certain times. I suspect the answer to all this is that these coordinates should reflect the raw (unflagged) data, and the user can consult the processing history for details of which parts of the data have been discarded... but be aware that this means the coordinate information is only a guideline as to what data the user might expect to find in a given image group, rather than being a definitive description. | open |
| 07 | I don't see any definition of how the image data itself is stored beyond as "N-dimensional array of doubles"? Is that flexible? Does it place reasonable limits on the dynamic range? Does HDF5 support an arbitrary precision floating point type? | open |
| 08 | What about image statistics? Integrated flux? Peak flux? Actual RMS? Thermal RMS limit? Where is information about the image resolution stored? Something like best fit and restored beam dimensions need to be kept someplace. There is one of these per image. Maybe in the image data attributes we store an array of values for these? | open |

*continued on next page*

| | | *continued from previous page* |
|---|---|---|
| Item | Description | Status |

## A.2. Future enhancements

The are a number of directions into which the current version of the format specification might be developed for future releases:

1. Frequency mapping information to track how the final frequency channels/bins in the image correspond to the frequencies in the original data from which the image was generated.
2. Mechanism to store beam-shape information.

# B. LOFAR Filename Convention

The LOFAR file naming convention is described in the document, `LOFAR-USG-ICD-005` [?]. Readers are encouraged to consult that document for specifics on LOFAR file naming conventions.

# C. Coordinates group examples

An in-depth description – including a number of examples – can be found in `LOFAR-USG-ICD-002` [?]. Readers are encouraged to consult that document for specifics on the storage of world coordinates information.

# Glossary of terms

**Az** Azimuth.

**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 tpyes of beams are available.

**BF** Beam-Formed data (time series structure).

**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 Casa-baseMeasurement 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 radio interferometric array. See `www.lofar.org` for details.

**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** 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** LOFAR Long Term Archive.

**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 CASA 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.

**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 [**?**, **?**, **?**].