LOFAR Data Format ICD
Visibility Data

Document ID: LOFAR-USG-ICD-007

Version 2.00.01

SVN Repository Revision: –revision–

K. Anderson

SVN Date: 2017-03-20

Contents

Change record

Version	Date	Sections	Description of changes
0.0	2010-04-09	All	Initial revision
0.01	2010-04-19	4	Refactor Sec. 4 == 4.1.1 4.1.2
0.1	2010-06-03	All	Initial commit
0.2	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` ("Representation of World Coordinates")
0.3	2010-06-06	??	Updated description of coordinates in order to keep up with changes 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.
2.00.01	2011-03-10	all	Maintain list of references through BibLaTeX database.

1 Introduction

1.1 Purpose and Scope

This document sets forth a formal data interface specification for LOFAR data products. The specification applies to data structures produced by various LOFAR processing pipelines that will be called LOFAR UV Visibility. This is a specification for LOFAR UV Visibility data products only and in no way implies, and should not be inferred as, a specification for any data structures the project may use during *in situ* processing by way of producing a final standard LOFAR UV Visibility file.

This document 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 UV Visibility file will be the data hosting structure for LOFAR UV Visibility data. It is therefore incumbent on the LOFAR project to define and describe both the structure of the LOFAR UV Visibility file format, and the various data types within the context of that format.

For the LOFAR project, a UV Visibility file 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 UV Visibility cube contents, conditions of observations, etc.. As an "all-in-one" wrapper, HDF5 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 UV Visibility file 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 files upon request, it is not entirely necessary that HDF5 header keywords match FITS keyword conventions in a LOFAR UV Visibility 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 UV Visibility files make use of *de facto* FITS standard keywords as much as possible.

For the purposes of further discussion regarding UV Visibility file adherence to FITS keyword standards, the *ESO Data Interface Control Document* (*see* § *References*), has been adopted as the FITS keyword model.

1.3 Applicable documents

Table **??** 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 [?]	TBB Time-Series Data	Digitized voltage output, as received by the individual LOFAR dipoles.
ICD-002 [?]	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 [?]	Beam-Formed Data	Hosting structure for LOFAR Beam-Formed data.
ICD-004 [?]	Radio Sky Image Cubes	Primary data product of the imaging pipeline.
ICD-005 [?]	File Naming Conventions	Conventions for the naming scheme applied to LOFAR standard data products.
ICD-006 [?]	Dynamic Spectrum Data	Hosting structure for dynamic spectrum data, i.e. intensity as function of time and frequency.

continued on next page

Applicable documents *continued from previous page*

Reference	Title	Description
ICD-007 [?]	Visibility Data	Hosting structure for LOFAR UV Visibility data, primary output of interferometer operations.
ICD-008 [?]	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 applicable to add the other ICDs.

2 Overview

This document is structured as follows: Section **??** 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. Section **??** will present the low-level specification for the data, including a description of the structure of LOFAR UV Visibility files, and the various group entities and sub-structures comprising these uv visibility data files, i.e. LOFAR group types, units, physical quantities. Finally, the LOFAR filename convention appears in Appendix **??**, and Coordinate group examples are present in Appendix **??**.

3 Organization of the data

3.1 High level LOFAR UV Visibility file structure

A LOFAR UV Visibility file will adhere to the following guidelines:

A LOFAR UV Visibility file will be defined within the context of the HDF5 file format. In an effort to minimize the hierarchical depth of the file structure, a UV Visibility file is designed to be a "flat" as possible, providing access to the necessary data without undue hierarchical tree crawling.

Therefore, the UV Visibility 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 "Vis Groups" (*see* § **??**). Each such group will contain UV visibility data and meta-data for a single sub-band of an observation.

A LOFAR UV Visibility file will then comprise an arbitrary, observation-dependent number of these Vis groups.

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

```
OBSERVATION/
OBSERVATION/SysLog
OBSERVATION/Vis001/Coordinates
OBSERVATION/Vis001/Coordinates/<DirectionCoord>
OBSERVATION/Vis001/Coordinates/<LinearCoord>
OBSERVATION/Vis001/Coordinates/<SpectralCoord>
OBSERVATION/Vis001/Coordinates/<TabularCoord>
OBSERVATION/Vis001/Coordinates/<PolarizationCoord>
OBSERVATION/Vis001/Data
OBSERVATION/Vis001/Source
OBSERVATION/Vis001/ProcessHist
...
```

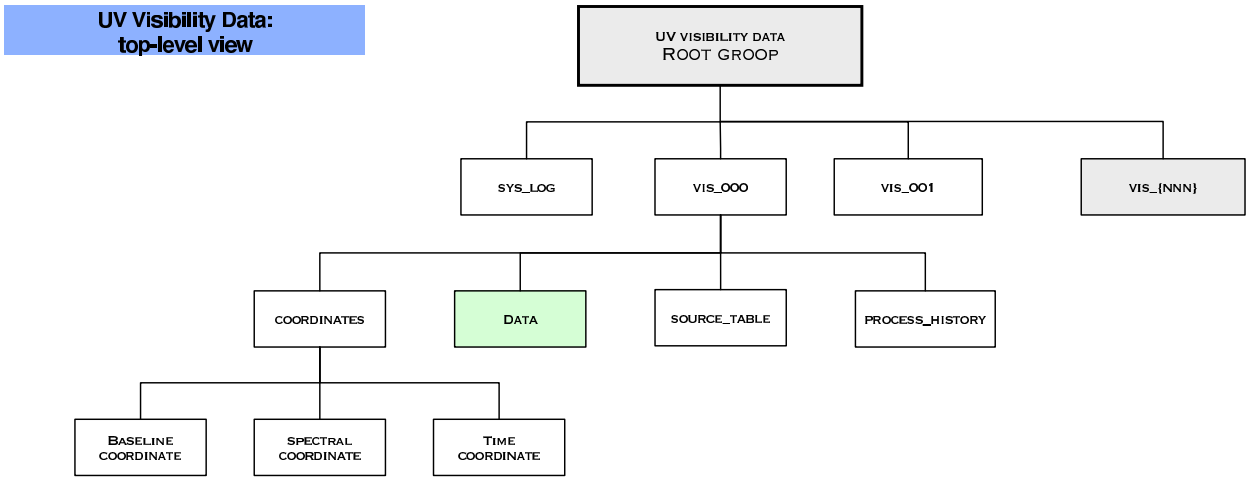

Figure 1: UV Visibility file structure

```
OBSERVATION/VisNNN/...
OBSERVATION/MashedVISImages/
...
```

3.2 Overview of UV Visibility Groups

A LOFAR UV Visibility file 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 UV Visibility file will contain an arbitrary, observation-dependent number of `Vis Groups` containing a `Coordinates Group`, a `Source Group`, and a `Processing History Group`, which contains pertinent logs and parameter sets of the relevent UV Visibilty sub-band.

These main building blocks of the UV Visibility HDF5 file are:

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. **??** 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. **??** for a detailed description.
3. **Vis Groups**. Each observation sub-band is stored as a seperate group within the file, containing its own set of four (4) sub-groups. Characteristics about each sub-band uv visibility are stored as Attributes in group headers. LOFAR imaging can produce up to 248 sub-bands, which will be the maximum number of uv visibilty groups possible in a LOFAR UV Visibility. Each uv visibility group will contain one Data group, which will in turn contain (1) dataset as an ndarray, along with associated attributes.
4. **Source Groups**. Each observation sub-band processing produces a file describing what is called the Local Sky Model (LSM). A **Source Group** will contain a table of the sources listed in a processing file, a `.skymodel` file. Attributes will described the columnar data.
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. **??** for a detailed description.

7. **UV Vis Sub-band Data arrays**. For each Vis Group, the subband uv visibility data are stored as ndarrays in the respective Data group – it is at this 4th hierarchical depth that the bulk of the data reside. The data storage options are still being investigated, in order to determine the maximum efficiency of data seeks and file I/O.

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 Visibility Data data, as well as providing hooks to all groups attached to the **File Root Group**.

This section will specify two set 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 UV visibility 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 visibility-specific.

In other words,

```
Root Attributes = Common LOFAR Attributes + Supplemental UV Vis Root Attributes.
```

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

4.1.1 Common LOFAR Attributes

Table 2 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 imaging, 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.

- 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 the Group Type table below, grouped by category.

- 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. **??**, 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 **??** below, or refer to [**?**]) of the LOFAR data file, indicates the kind of LOFAR data contained.
- 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'.

General LOFAR Group	Value	Description
Root	'Root'	Top-level LOFAR group type
System Log	'SysLog'	System log files, parsets
Average Vis	'MashedUVVis'	Summed uv visibility
Vis	'Vis'	Vis group
Vis Group Subgroups	**Value**	**Description**
Data group	'Uvvis'	This is a UV Visibility Data group
Source group	'Source'	This is a Source List group
Processing History group	'ProcessHist'	This is a Processing History group
Masks group*	'Masks'	This is a Masks group
Coordinates Group	'Coordinates'	This is a Coordinates group
Coordinates Group Subgroups	**Value**	**Description**
Direction coord group	'DirectionCoord'	This is a direction coord group
Linear coord group	'LinearCoord'	This is a linear coord group
Spectral coord group	'SpectralCoord'	This is a Spectral coordinate group
Tabular coord group	'TabularCoord'	This is a tabular coord group
Polarization coord group	'PolarizationCoord'	This is a Polarization coordinate group
**Proposed groups under advisement.*		

Table 2: LOFAR UV Visibility Group Types

- 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) using NNNNNN.NNNNNNN format,
  - International Atomic Time (OBSERVATION_START_TAI) using yyyy-mm-ddThh:mm:ss.sssssssss format and
  - Coordinated Universal Time (OBSERVATION_START_UTC) using yyyy-mm-ddThh:mm:ss.sssssssssZ format.
- The observation's end time is listed in the following formats:
  - Modified Julian Day (OBSERVATION_END_MJD) using NNNNNN.NNNNNNN format,
  - International Atomic Time (OBSERVATION_END_TAI) using yyyy-mm-ddThh:mm:ss.sssssssss format and
  - Coordinated Universal Time (OBSERVATION_END_UTC) using yyyy-mm-ddThh:mm:ss.sssssssssZ format.
- 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:

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

  Given the possibilities of rather non-regular coverage in frequency space, ν_{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.

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	double	—	Observation start date (MJD)
OBSERVATION_START_TAI	string	—	Observation start date (TAI)
OBSERVATION_START_UTC	string	—	Observation start date (UTC)
OBSERVATION_END_MJD	double	—	Observation end date (MJD)
OBSERVATION_END_TAI	string	—	Observation end date (TAI)
OBSERVATION_END_UTC	string	—	Observation end date (UTC)
OBSERVATION_NOF_STATIONS	int	—	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
ICD_NUMBER	string	—	Interface Control Document number
ICD_VERSION	string	—	Interface Control Document version/issue number
NOTES	string	—	Notes or comments

Table 3: 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.
Instrument Model	'inst'	Parameters describing gain and other instrument characteristics for calibation.
Sky Model	'lsm'	List of sources, either point sources or shapelets.

Table 4: Overview of standard LOFAR data products and the corresponding file type attribute value.

- 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 **??** 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 5: 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 **??** 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 6: 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.
- `ICD_NUMBER` and `ICD_VERSION` list name/number and version/issue of the Interface Control Document (ICD) to which the data abide by.
- The `NOTES` attributes acts as generic area for notes and comments.

4.1.2 Supplemental UV Visibility Root Attributes

The root group of a LOFAR UV Visibility file will comprise header attributes, various subgroups as indicated above, and appropriate pointers to root-level an arbitrary number of `Vis` sub-groups, wherein each Vis group comprises the relevent data and meta-data for a single sub-band of an observation.

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 **??** presents additional root group attributes for a LOFAR UV Visibility file that do not appear in the LOFAR common metadata table.[1]

Field/Keyword	Type	Value	Description
`VISGROUPS`	bool	'true'	File has uv visibility subgroups
`NOF_IMAGES`	int		`N` Vis groups in this file
`RA_TARG`	float		`RA` of `TARGET` (at LOFAR core)
`DEC_TARG`	float		`Dec` of `TARGET` (at LOFAR core)
`ORIGFILE`	string		Input data file (MS?)

Table 7: Additional Root group attributes, LOFAR UV Visibility

4.2 The Vis Group

The `Vis` group will be an HDF5 group serving as a container for the four sub groups described below. As far as is posssible, an `Vis` group is designed to be a complete and self-contained package of uv visibility data. It will contain relevant data and metadata for a particular processed sub-band of a LOFAR observations. 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-compatable form.

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

- A `Coordinates` group that will contain one or more subgroups of kinds `LinearCoord, TabularCoord, SpectralCoord, DirectionCoord, PolarizationCoord`, that will describe various axes of the associated dataset.
- A `Data` group that will contain a dataset array.
- A `Source` group that will be a tabular representation of a Local Sky Model.
- A `ProcessHist` group, which will be a meta-data container holding various processing products such as log files, parameter sets, RFI mitigation tables, etc.

Figure 2 illustrates the form of an Vis group in a LOFAR UV Visibility file. The table of Vis 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 Vis group metadata (*see* **??**, *"Coordinates group"*)

Field/Keyword	Type	Value	Description
GROUPTYPE	string	'Vis'	LOFAR group type
COORDINATES	string		name of 'Coordinates' subgroup
DATAGROUP	string		name of 'Data' subgroup
SOURCEGROUP	string		name of 'Source' subgroup
PROCESS_HISTORY	string		name of 'ProcessHist' subgroup

Table 8: Vis group Attributes

4.3 The Coordinates Group

Coordinate information within a LOFAR UV Visibility file 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 will be a subgroup of an `Vis` 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, linear, tabular, spectral, polarization`.

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. **??** 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.

The attributes, as presented in Table **??**, summarize the overall characteristics of the set of coordinates collected within this 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.

4.3.1 Linear coordinate

As already indicated by the name, this group encodes the properties of a simple linear coordinate (or a number thereof, as multiple axes are permitted). The attributes storing the actual coordinate parameters are listed in Tab. **??** below.

Field/Keyword	Type	Value	Description
GROUPTYPE	string	'LinearCoord'	Group type descriptor
COORDINATE_TYPE	string	'Linear'	Coordinate Type descriptor
STORAGE_TYPE	string	'Linear'	Descriptor for the underlying storage type for this coordinate
NOF_AXES	int	N	Number of coordinate axes
AXIS_NAMES	array<string,1>	$[name_0, .., name_N]$	World axis names
AXIS_UNITS	array<string,1>	$[unit_0, .., unit_N]$	Physical units along each coordinate axis.
REFERENCE_VALUE	array<double,1>	$[val_0, .., val_N]$	Coordinate value at the reference point
REFERENCE_PIXEL	array<double,1>	$[pix_0, .., pix_N]$	Array location of the reference point in pixels.
INCREMENT	array<double,1>	$[incr_0, .., incr_N]$	Coordinate increment at reference point.
PC	array<double,1>	$[p_{00}, pc_{01}, .., p_{0N}, .., p_{NN}]$	Non-singular square matrix, for the transformation from intermediate pixel coordinates to intermediate world coordinates.

Table 10: Keywords decribing a Linear Coordinate.

4.3.2 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 ν, λ, 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 **??** below lists the various spectral quantities and their respective encoding as an attribute; the symbols λ_0 and ν_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 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.

Attribute	FITS Code	Name	Symbol	Associate variable	Default units
Frequency	FREQ	Frequency	ν	ν	Hz
Energy	ENER	Energy	E	ν	J
Wavenumber	WAVN	Wavenumber	κ	ν	m^{-1}
VelocityRadio	VRAD	Radio velocity	V	ν	m s^{-1}
VelocityOptical	VOPT	Optical velocity	Z	λ	m s^{-1}
VelocityAppRadial	VELO	Apparent radial velocity	v	v	m s^{-1}
Redshift	ZOPT	Redshift	z	λ	–
WavelengthVacuum	FREQ	Vacuum wavelength	λ	λ	m
WavelengthAir	AWAV	Air wavelength	λ_a	λ_a	m
BetaFactor	BETA	Beta factor v/c	β	v	–

Table 11: 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.

4.3.3 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. **??** below.

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. **??** 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, ν_0
REST_FREQUENCY_UNIT	string	'Hz'	Physical units within which the rest frequency is given
REST_WAVELENGTH	double		Rest wavelength, λ_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_{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>	$[p_0, .., p_{N_{\mathrm{Pixels}}}]$	Tabulated values along the pixel axis.
AXIS_VALUES_WORLD	array<double,1>	$[w_0, .., w_{N_{\mathrm{Pixels}}}]$	Tabulated values along the world axis.

Table 12: Keywords decribing a Spectral Coordinate; attributes within the first segment of the table will be present independent of the specific storage method.

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_{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 13: Keywords decribing a Polarization Coordinate.

```
.
'- COORDINATE_{N}
   |- GROUPTYPE              Group       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>
   |- AXIS_LENGTH            Attr.       int
   |- AXIS_VALUES_PIXEL      Attr.       array<double,1>
   '- AXIS_VALUES_WORLD      Attr.       array<double,1>
```

Listing 1: Structure of the tabular coordinate group.

4.3.4 Tabular coordinate

Comment:
Add paragraph outlining motivation and need for tabular coordinate: representation of non-linear (tabulated) 1-dimensional coordinate axis.

- GROUPTYPE is the group type descriptor with the fixed value 'TabularCoord'.
- COORDINATE_TYPE is the is the descriptor for the coordinate type, of value 'Tabular'.
- STORAGE_TYPE is the descriptor for the underlying storage type for this coordinate, of value 'Tabular'.
- NOF_AXES is the number of coordinate axes; keep in mind that a coordinate can consist of multiple axes.
- AXIS_NAMES are the world axis names connected with the coordinate axes, e.g.

  AXIS_NAME=['Distance']
  AXIS_NAME=['Time']

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 14: Recognized values for the Polarization component parameter.

- `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 [**?**].
- `AXIS_VALUES_PIXEL` are the tabulated values of pixel coordinates.
- `AXIS_VALUES_WORLD` are the tabulated values of world coordinates.

4.4 The Data Group

A `Data` group will most often be a subgroup of an `Vis` group container and consist of an HDF5 "dataset," which, as defined in the HDF5 documentation (*see* § *References*), is "stored in two parts: a header and a data array." However, with the adoption of a so-called "Coordinates group," which contains all the relevent pointing, projection, and unitary information, scale and unit metadata, `Data` group attributes will be limited.

LOFAR UV Visibility files will limit attributes to nominal keyword-value pairs as much as possible, with a thought toward potential future user requests for FITS format images. See § **??** "The Coordinates group," for a detailed specification of `Data` group header attributes.

The dataset array will (usually) be a ndarray data structure, as can be created by the Python *numarray/numpy* packages. The nominal dimensionality of a `Data` group's dataset will be 4 (`NAXIS=4`), wherein the image cube (or cubes) will be defined in (C-type order) *Polarization*, *Frequency*, *Dec*, *RA*,

Field/Keyword	Type	Value	Description
GROUPTYPE	string	'TabularCoord'	Group type descriptor
COORDINATE_TYPE	string	'Tabular'	Coordinate Type descriptor
STORAGE_TYPE	string	'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>	$[name_0]$	World axis names
AXIS_UNITS	array<string,1>	$[unit_0]$	Physical units along each coordinate axis.
AXIS_LENGTH	int	N_{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<T,1>	$[p_0, .., p_{N_{\mathrm{Length}}}]$	Tabulated values along the pixel axis; depending on the quantity represented T={double,int}.
AXIS_VALUES_WORLD	array<T,1>	$[w_0, .., w_{N_{\mathrm{Length}}}]$	Tabulated values along the world axis; depending on the quantity represented T={double,string}.

Table 15: Keywords decribing a Tabular Coordinate.

Field/Keyword	H5Type	Type	Value	Description
GROUPTYPE	Attribute	string	'Data'	Group type descriptor
DATASET	Attribute	bool	'true'	the group contains a data array
WCSINFO	Attribute	string	'/Coordinates'	hdf5 path to Vis group WCS data

Vis	Quantity	Axes	Units
Sky image	$I(p, \nu, \mathrm{Dec}, \mathrm{RA})$	Pol/Freq/Dir/Dir	.. /Hz/deg/deg
RM cube	$RM(p, \phi, \mathrm{Dec}, \mathrm{RA})$	Pol/Faraday Rot/Dir./Dir.	.. /rad m^{-2}/deg/deg
RM map	$RM(\mathrm{Dec}, \mathrm{RA})$	Dir./Dir.	/deg/deg
CR image	$I(p, \nu, r, \mathrm{ElAz})$	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

4.5 The Source Group

The Source group in a UV Visibility file will be a table of sources and their associated parameters. The Source group header will specify the fields (columns) of the table, the number of sources in the table (rows). See Table ?? for the specification of Source group attributes for a LOFAR UV Visibility file.

4.6 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 UV Visibility file HDF5 groups and subgroups, the Processing History group will be an HDF5 group, as a subgroup of an Vis 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	Type	Value	Description
GROUPTYPE	string	'Source'	UV Visibility group type
DATASET	string	'Source List'	These data are a local sky model
NAXIS	int	2	Number of data axes
NAXIS1	string	'Fields'	Axis of the data fields
NAXIS2	string	'Source'	Axis of the source rows.
NSOURCE	int		Number of data rows/sources
FIELD1	float		RA (at LOFAR core)
FIELD2	float		Dec (at LOFAR core)
FIELD3	float		Peak Flux
FIELD4	float		Integrated Flux
FIELD5	float		Gaussian semi-major axis
FIELD6	float		Gaussian semi-minor axis
FIELD7	float		Position angle

Table 16: Attributes of a Source group.

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 17: Attributes of a Processing History group.

5 Interfaces

—/—

5.1 Interface requirements

—/—

5.2 Relation to other workpackages

—/—

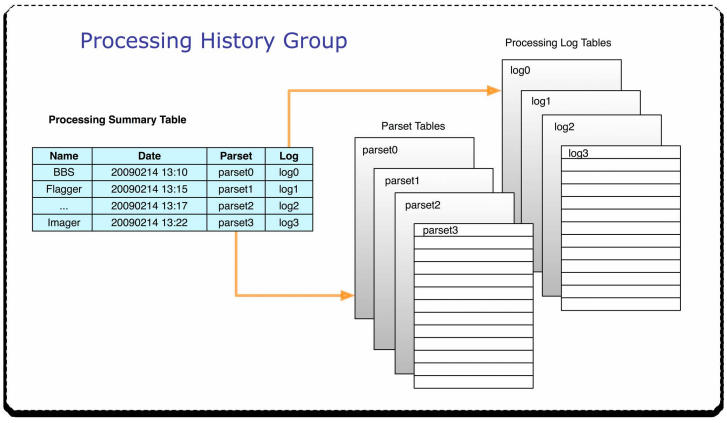

Figure 2: The processing history group, nested tabulation

A Discussion & open questions

A.1 Open questions/Issues

1. Do we really need a full description of all coordinate types within this document? Since this information written down in detail in LOFAR-USG-ICD-002, we rather might want to restrict ourselves here to the specific metadata used for Visibility data.

A.2 Future enhancements

—/—

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.

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