LOFAR Data Format ICD
Rotation Measure Synthesis Cubes

Document ID: LOFAR-USG-ICD-008

Version 0.07.07

SVN Repository Revision: 9370

J. Anderson, L. Bähren, M. Bell, T. Riller, K. Anderson

SVN Date: 2012-01-10

Contents

Change record

Version	Date	Sections	Reason
0.01.00	2010-06-02	All	Initial version
0.02.00	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")
0.03.00	2010-07-15	All	Incorporated James Anderson's comments from the forum.
0.04.00	2010-12-07	All	Using LaTeX package `hyperref` for references, enabling better navigation through the document and access to external resources. Put back in usage of `a4wide` package, in order to make the textbox consistent with that of the other ICDs.
0.05.00	2011-01-04	All **??**, **??**	Review, cleanup, and overhaul, new diagram illustrating the latest design.
0.05.01	2011-01-17	—	Minor revision, activation of svn keywords.
0.05.02	2011-01-18	**??**, **??**	Correction of fontsize. Adjusted attribute names.
0.06.00	2011-01-19	All	Removed several poorly defined and at the moment extraneous groups. Now focusing on the basic requirements for the description of a RMSC. Added a polarization coordinate group.
0.07.00	2011-02-01	§**??**	move and insert image table, descr. text
0.07.01	2011-03-10	all	Maintain list of references through BibLaTeX database.
0.07.02	2011-03-15	all	Fixed some typos and changed all group and attribute names so that they are UPPER_CASE rather than CamelCase.
0.07.03	2011-07-06	all	Matching up group type attributes and notation; consolidation of labels to refer to standard sections and tables.
0.07.04	2011-09-20	**??**	Update of Fig. **??**
0.07.05	2011-09-28	all	Change data types of attributes: `float` → `double`.
0.07.06	2011-10-25	all	Use names `PROCESS_HISTORY` and `SYS_LOG` consistently throughout the document. Change data type of attributes: `bool` → `unsigned int`.
0.07.07	2012-01-10	all	Change data type of attributes: `unsigned int` → `bool`; added input tex file on types to the section just before Acknowledgements. Changed the svnInfoRevision to svnInfoMaxRevision, in order to take the sub-tex file changes into account for the latex compile.

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

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

where

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

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

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

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

Notation.

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

Acknowledgements

The document was originally developed by K. Anderson in collaboration with L. Bähren.

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 *Rotation Measure Synthesis Cubes* (henceforth RMSC). This is a specification for RMSC 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 Rotation Measure Synthesis Cube 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 RMSC file will be the data hosting structure for LOFAR RM synthesis output data. It is therefore incumbent on the LOFAR project to define and describe the structure of the LOFAR RMSC file format, and how the various data types are defined and described within the context of that format.

While image hypercubes are the primary data product of the RM synthesis 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 RMSC 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 RMSC 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.

For the purposes of further discussion regarding RMSC file adherence to FITS keyword standards, the *ESO Data Interface Control Document*, 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.
`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.

continued on next page

Applicable documents *continued from previous page*		
REFERENCE	TITLE	DESCRIPTION

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

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 ("CR image" in Table ??), etc., all have different dimensions and coordinate types. Table ?? illustrates the various image data array dimensions that LOFAR imaging may produce.

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

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 "RMSC" (Rotation Measure Synthesis Cube). The dataset array in a RMImage Cube will be a `ndarray` data structure, as can be created by the C-based *numarray/numpy* Python packages. The Rotation Measure Synthesis Cube is designed to store the 3D polarized intensity data produced from the LOFAR RM synthesis pipeline. The nominal dimensionality of a RMSC `DATA` group's dataset will be 4 (`NAXIS=4`), wherein the image cube (or cubes) will be defined in (C-type order) *Polarization*, *Faraday Depth*, *Dec*, *RA*, as shown in Table ??

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 RM Synthesis Cube files, and the various group entities and sub-structures comprising these RMSC 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 RM Synthesis Cube file structure

A LOFAR Rotation Measure Synthesis Cube (RMSC) file will adhere to the following guidelines:

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

Therefore, the RMSC 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 a "System Log," as well as a single "RMIMAGE" group (*see* § **??**), where an `RMIMAGE` group will contain data and meta-data for a specified set of Faraday depths.

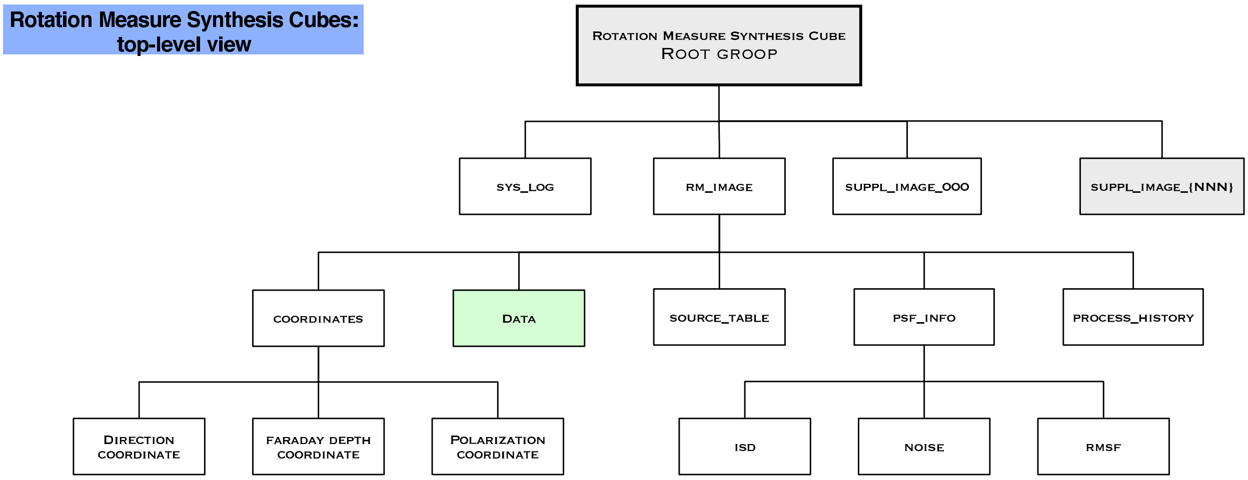

Figure 1: Rotation Measure Synthesis Cube (RMSC) file structure

Unlike the SkyImage Cube, a LOFAR RMSC file will *only require a single RMImage group*. Separation in to "sub-bands", each having different properties, is unnecessary.

The structure of an RMSC file can be represented through HDF5 as a POSIX-style hierarchy[1]

```
OBSERVATION/
OBSERVATION/SYS_LOG
OBSERVATION/RMIMAGE/
OBSERVATION/RMIMAGE/COORDINATES
OBSERVATION/RMIMAGE/COORDINATES/<DIRECTION_COORD>
OBSERVATION/RMIMAGE/COORDINATES/<FARADAY_DEPTH_COORD>
OBSERVATION/RMIMAGE/COORDINATES/<POLARIZATION_COORD>
OBSERVATION/RMIMAGE/DATA
OBSERVATION/RMIMAGE/SOURCE_TABLE
OBSERVATION/RMIMAGE/PROCESS_HISTORY
OBSERVATION/RMIMAGE/PSF_INFO
...
```

3.2 Overview of RMSC Groups

A LOFAR RMSC file will then comprise `SYS_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 RMSC file will contain a *single* `RMIMAGE` group containing a `DATA` group, a `COORDINATES` group, a `SOURCE_TABLE` table, a `PSF_INFO` group and a `PROCESS_HISTORY` group, which contains pertinent logs and parameter sets of the relevant image.

These main building blocks of the RMSC 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.

[1] < ... > Indicates optional or potiential subgroups, where only one coordinate subgroup of any type shall be present in a LOFAR RMSC file.

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. **RMImage Group**. The `RMIMAGE` group contains its own set of five (5) sub-groups. Characteristics about the Rotation Measure Image are stored as Attributes in group headers. The `RMIMAGE` group will contain one (1) `DATA` group, which will in turn contain one (1) dataset as an ndarray, along with associated attributes.
4. **Source Tables**. A `SOURCE_TABLE` Table will contain a table of the sources listed in a processing file, a `.skymodel` file. Attributes will describe the columnar data. This table will also contain the model components obtained from one of several possible types of imaging routines (e.g. CLEAN, Wavelet, etc.).
5. **Coordinates Groups** (`COORDINATES`). The `RMIMAGE` group contains one `COORDINATES` group, which stores the relevant world coordinate conversions. The Coordinates group *must contain at least one subgroup* of the kind direction and Faraday depth. Multiple coordinate subgroups are permissible as needed, but only one of any type can be present in a LOFAR RMSC file. The `COORDINATE` group will contain 3 or $3 \times N$ Coordinate sub-groups.
6. **Processing History Groups** (`PROCESS_HISTORY`) can be found on the `RMIMAGE` 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. **Data Group/Data arrays** (`DATA`). The rotation measure synthesis image data are stored as ndarrays in the respective `DATA` group - it is at this 3rd hierarchical depth where the image data reside. The data storage options are still being investigated, in order to determine the maximum efficiency of data seeks and file I/O.
8. **PSF Info Groups**. The `PSF_INFO` group is a subgroup of the `RMIMAGE` group and contains information about the **point spread function** (PSF) in both the sky plane and along the Faraday depth axis. This includes the basic dimensions of the PSF, the full 1-D PSF in Farday depth, and the list of frequencies and channel widths from which the Faraday depth PSF has been derived.

4 Detailed Data Specification

4.1 The Root Group (ROOT)

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

This section will specify two sets of attributes that will appear in the `ROOT` group: a set of Common LOFAR Attributes (CLA) that will be common to all LOFAR science data products, and a set of attributes that are specific to LOFAR RMSC 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 `RMIMAGE`-specific. In other words,

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

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

4.1.1 Common LOFAR Attributes

This section will specify a set of attributes that will be common to LOFAR science data products. These "Common LOFAR Attributes" will appear as attributes at the root level of all LOFAR data files. *All* LOFAR data products, including RMSC *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 ?? 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 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 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'.
- 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:

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

  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.

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.

General LOFAR Group	**Value**	**Description**
Root	'Root'	Top-level LOFAR group type
System Log	'SysLog'	System log files, parsets
SIG*	'SIG'	Supplemental Image Group
RMImage	'RMImage'	RMImage group
RMImage Group Subgroups	**Value**	**Description**
Data group	'rmscData'	RMSC Data group (dataset)
Source table	'SourceTable'	Source List table
Processing History group	'ProcessHist'	Processing History group
Masks group*	'Masks'	Masks group
PSF Info Group	'PSFInfo'	PSF information group
Coordinates Group	'Coordinates'	Coordinates group
Coordinates Group Subgroups	**Value**	**Description**
Direction coord group	'DirectionCoord'	Direction coordinate group
Faraday depth coord group	'FaradayDepthCoord'	Faraday depth coordinate group
Polarization coord group	'PolarizationCoord'	This is a Polarization coord group
**Proposed groups under advisement.*		

Table 7: LOFAR RMSC Group Types

4.1.2 Supplemental RMSC Root Attributes

The root group of a LOFAR RMSC file will comprise header attributes, various subgroups as indicated above, and appropriate pointers to the root-level of a single RMIMAGE sub-group, which contains the relevant data and meta-data for a specified set of Faraday depths.

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 RMSC file that do not appear in the LOFAR common metadata table.[2]

Field/Keyword	Type	Value	Description
IMGROUPS	bool	'true'	File has RMImage subgroups
NOF_IMAGES	int	'1'	1 RMImage group in this file
TARGET_RA	double		RA of TARGET (at LOFAR core)
TARGET_DEC	double		Dec of TARGET (at LOFAR core)
INPUT_FILE	string		Input data file (SkyImage Cube)

Table 8: Additional Root group attributes, LOFAR RMSC

4.2 The System Logs Group (SYS_LOG)

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

4.3 The RMImage Group

The RMIMAGE group will be an HDF5 group serving as a container for the five sub groups described below. A RMIMAGE group is designed to be as complete and self-contained a Rotation Measure Synthesis cube as possible. It will contain relevant data and metadata for a processed cube of polarized intensity as a function of Faraday depth (known as the Faraday dispersion function). 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.

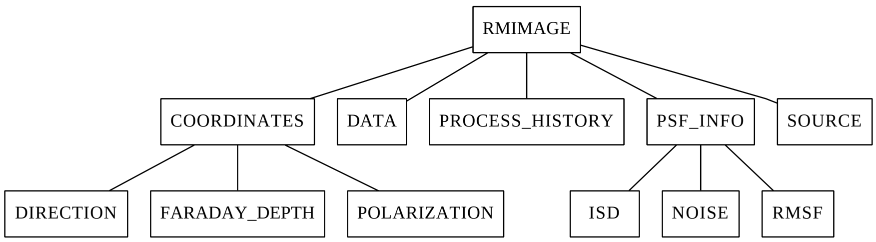

Figure 2: RMIMAGE Group structure

A RMIMAGE group will comprise five sub groups. These five groups, two hierarchical levels below the root group in a LOFAR RMSC *(Fig. 1)*, will be

- A COORDINATES group that will contain FARADAY_DEPTH_COORD, DIRECTION_COORD and POLARIZATION_COORD subgroups that describe the axes of the associated dataset.
- A DATA group that will contain a dataset array.
- A SOURCE_TABLE table that will be a tabular representation of a Local Sky Model as well as any image model component (e.g. CLEAN) information.

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

- 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..
- A PSF_INFO group, which contains information about the point spread function in both the image plane (i.e. the CLEAN beam) and along the Faraday depth axis (i.e. the RM Spread Function). The full RM spread function may be stored here (in 1-D for the time being).

The table of RMIMAGE group attributes, table **??**, is notably sparse here and the reader must bear in mind that the Coordinates groups will contain most of the rest of the relevant Image group metadata (*see* **??**, *"Coordinates group"*)

Field/Keyword	H5Type	Type	Value	Description
GROUPTYPE	Attr.	string	'RMIMAGE'	LOFAR group type
RMIMAGE	Attr.	string	'RMImage_{NNN}'	Name of the image group.
COORDINATES	Attr.	string	—	name of 'Coordinates' subgroup
DATA	Attr.	string	—	name of 'Data' subgroup
SOURCE_TABLE	Attr.	string	—	name of 'Source' subgroup
PROCESS_HIST	Attr.	string	—	name of 'ProcessHist' subgroup
PSF_INFO	Attr.	string	—	name of 'PSFInfo' subgroup
COORDINATES	Group			
DATA	Dataset			
SOURCE_TABLE	Dataset			
PSF_INFO	Group			
PROCESS_HISTORY	Group			

Table 9: Components of the RMImage group.

4.4 The Coordinates Group (COORDINATES)

Coordinate information within a LOFAR RMSC 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 RMIMAGE 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, faraday depth, 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 10: Components of a Coordinates group.

The attributes 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. For RMSC files, NOF_COORDINATES will be 2, with one coordinate of type 'DirectionCoord', and one coordinate of type 'FaradayDepthCoord'.
- 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.

Comment:
The description of the coordinates groups needs some major cleaning up: a) How many axes and coordinates are there for a RM Synthesis Cube? b) What is the ordering of the coordinate axes? c) Some of the description below appears to come from the Sky Cube ICD and might not be required here.
Mike: There are three coordinate axes - Dec., RA, FD. This should be the order. An optional spectral coordinate would contain information about the input spectral distribution (although this ideally belongs in the PSF information group if possible since it determines the RMSF and is not needed for the definition of the RM cube).

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. The attributes storing the actual coordinate parameters are listed in Tab. **??** below.

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. **??** below.

4.4.2 Polarization coordinate

The data contained within the RMSC is inherently complex. Rather than store a cube of complex values, the LOFAR RMSC stores the real and imaginary parts of the polarization data separately and so a 'Polarization' coordinate is required. For a RMSC, the Polarization coordinate will be of length 2 and include only Stokes Q and U. Following the normal conventions, the real part of the polarized intensity is stored in Stokes Q and the imaginary part in Stokes U.

- 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'.

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, ϕ_p
LATPOLE	double		Native latitude of the celestial pole, θ_p.

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

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

- 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:

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.

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.

4.4.3 Faraday depth coordinate

A Faraday depth coordinate can be a defined as either a linear or a tabular coordinate. This coordinate will typically be linear since the RMSC is produced using a discrete or fast Fourier transform, but because some circumstances (i.e. omission of data in empty regions along the Faraday depth axis) will require tabular coordinates the option is provided.

The Faraday depth coordinate information can be stored in either a set of parameters describing a linear axis (REFERENCE_PIXEL, REFERENCE_VALUE, INCREMENT, PC) or a tabulated axis (AXIS_VALUES_PIXEL, AXIS_VALUES_WORLD). The choice of encoding the FD coordinate information is reflected in the STORAGE_TYPE attribute: if set *Linear* the coordinate axis is expected to be linear, thereby using the first set of attributes. If set *Tabular*, the values along the coordinate axis are expected to be tabulated, thereby using the second set of attributes.

Comment:
Include a polarization coordinate group description

4.5 The Data Group (DATA)

A Data group is a subgroup of an RMIMAGE group container and consist of an HDF5 "dataset," which, as defined in the HDF5 documentation [?, ?, ?], is "stored in two parts: a header and a data array." However,

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.

the adoption of the `'COORDINATES'` structure, which contains all the relevant pointing, projection, and unitary information, scale and unit metadata, means that `DATA` group attributes will be necessarily limited.

LOFAR RMSC 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 WCS information and other relevant observation attributes.

The dataset array will (usually) be a complex 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 RM synthesis cube (or cubes) will be defined in (C-type order) *Pol*, *Dec*, *RA*, *Faraday depth*.

4.6 The Source Table (SOURCE_TABLE)

The `SOURCE_TABLE` table in a RMSC file will collect sources and an their associated parameters, as extracted from the Local Sky Model (LSM) or from image modeling routines such as CLEAN. The `SOURCE_TABLE` table header will specify the fields (columns) of the table, the number of sources in the table (rows). See Table **??** for the specification of `SOURCE_TABLE` group attributes for a LOFAR RMSC file.

> **Comment:**
> I have some questions about how this should be handled... Do we need to define several standard types of column headings here? Also, should we keep separate count of sources from the LSM and from the image models? I think that could be helpful... Also note that the Image cube also needs to have the capacity for storage of CLEAN components. – Mike

- `GROUPTYPE` – Group type of this structure, `SourceTable`
- `NOF_SOURCES` – Number of sources listed/collected in the source table.
- `COLUMN_NAMES` – Name of the table columns, see below.

Field/Keyword	Type	Value	Description
GROUPTYPE	string		Group type descriptor, 'FaradayDepthCoord'
COORDINATE_TYPE	string		Coordinate Type descriptor, 'FaradayDepth'
STORAGE_TYPE	string		Descriptor for the underlying storage type for this coordinate, e.g. Linear or Tabular
NOF_AXES	int	'1'	N of coordinate axes, always 1
AXIS_NAMES	array<string,1>		World axis names
AXIS_UNITS	array<string,1>		World axis units
REFERENCE_VALUE	array<double,1>		Reference value
REFERENCE_PIXEL	array<double,1>		Reference pixel
INCREMENT	array<double,1>		Coordinate increment
PC	array<double,1>		
AXIS_VALUES_PIXEL	array<double,1>		Reference pixels
AXIS_VALUES_WORLD	array<double,1>		Reference values

Table 15: Keywords decribing a Faraday Depth 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	'Data'	Group type descriptor
DATASET	bool	'true'	the group contains a data array
WCSINFO	string	'/Coordinates'	hdf5 path to RMImage group WCS data
DATAUNITS	string	Jy rad^{-1} m^{2}	Units of the main data cube
GLOBAL_ERROR_VALUE	double	—	The estimated RMS error in either the real or imaginary coordiantes, averaged over the entire cube

Table 16: The attributes of the DATA group.

- 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. **??**.
- 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.

Comment:
The entries in the source table – it actually is not a group – need to be cleaned up. a) Do we need to track uncertainties as well, i.e. errors on the source position? b) 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?
Mike: a) maybe b) it would be best if all entries had the same units. I don't see why they shouldn't. c) We will need some kind of spectral index as well, both for peak and integrated flux I would think. maybe a spectral curvature measure as well since i would imagine many source spectra will be turning over at LOFAR frequencies.

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.
COLUMN_NAMES	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.
FD_UNITS	string		Physical units of Faraday Depth – rad m^{-2} – within which the source position is recorded.

Table 17: Attributes of the Source table; attributes visible at this level will be shared across all entries within the table.

Column/Keyword	Type	Value	Description
NUMBER	int	—	Running index of the table entry.
NAME	string	—	Name of the source
ORIGIN	string	—	Origin of the source model (e.g. LSM, CLEAN)
RA	double	—	RA position of the source.
DEC	double	—	Dec position of the source.
FD	double	—	Faraday depth 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 used for 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 18: Columns of the source table.

4.7 PSF Information Group

The `PSF_INFO` group includes the basic PSF information i.e. the FWHM of the beam in the sky plane and of the RMSF main peak. It also contains the complete RMSF and the input spectral distribution.

FIELD/KEYWORD	TYPE	VALUE	DESCRIPTION
GROUPTYPE	string	'PSF_INFO'	LOFAR group type
MSB_MAJOR_AXIS	double	—	Mean synthesized beamsize semimajor axis, in arcseconds
MSB_MINOR_AXIS	double	—	Mean synthesized beamsize semiminor axis, in arcseconds
MSB_POSITION_ANGLE	double	—	Mean synthesized beamsize position angle
MSB_POSITION_ANGLE_UNIT	string	'deg'	Physical units for the synthesized beamsize position angle
RES_MAJOR_AXIS	double	—	Restoration beamsize semimajor axis, in arcseconds
RES_MINOR_AXIS	double	—	Restoration beamsize semiminor axis, in arcseconds
RES_POSITION_ANGLE	double	—	Restoration beamsize position angle, in degrees
RES_POSITION_ANGLE_UNIT	string	'deg'	Physical units for the synthesized beamsize position angle
MEAN_ERR	double	—	Mean error in the Faraday depth brightness estimate
RMSF_FWHM	double	—	Gaussian fit FWHM of the RMSF, in rad m^{-2}
RMSF_RES_FWHM	double	—	Gaussian FWHM of the restoration RMSF, in rad m^{-2}
ISD	array<double,2>	—	The frequencies and bandwidths from which the RMSF has been derived
ISD_UNITS	string	'MHz'	The units in which the ISD and bandwidths are defined.
DELTA_FD_RMSF	double	—	The cell size for the FD axis which describes the RMSF
NUM_FD_RMSF	int	—	The number of cells along the FD axis which describes the RMSF
UNITS_FD_RMSF	string	'rad/m/m'	Units of the FD axis which describes the RMSF
RMSF	array<double,2>	—	An array containing the real and imaginary parts of the RMSF

Table 19: Attributes of the PSF subgroup.

The full RMSF is stored as an 2 column array of numbers with the first column being the real part of the RMSF (Stokes Q) and the second column being the imaginary part (Stokes U). An RMSC generically has a 3-D point spread function that is position dependent. In future implementations, the ability to store several RMSFs corresponding to different reference locations will provided.

Comment:
I think I will need help here. We need to store the RMSF as well as some information about the FD axis on which it is defined (shouldn't simply be the image FD axis...) I see no reason why we shouldn't force this to be a linear axis type, and it will always be centered on zero. So we just store 3 values to describe the axis: deltaFD, numFD, and unitsFD. Then we just store the RMSF in an Nx2 array. ISD

Though a spectral coordinate is not needed for the definition of the RMSC, a record of the spectral properties of the input sky image data is kept for diagnostic and preservation purposes.

4.8 The Processing History Group (PROCESS_HISTORY)

The data definition for the PROCESS_HISTORY group (PROCESS_HISTORY, as it will appear in LOFAR data products) 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 DP3 log files, processing parameters sets, RFI mitigation tables, etc. In fact, and due the wide-ranging data types and free-form ASCII format of many log files that the PROCESS_HISTORY group may encompass, this group will be a catch-all envelop encapsulating information about all steps of processing should the user need such information. And it is because of this free-form nature of the meta-data that it is very difficult to define a header describing attached data when it is not yet know just what those data may include. An attempt has been made to provide by example how this will or should appear in the PROCESS_HISTORY group header.

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 20: Attributes of a Processing History group.

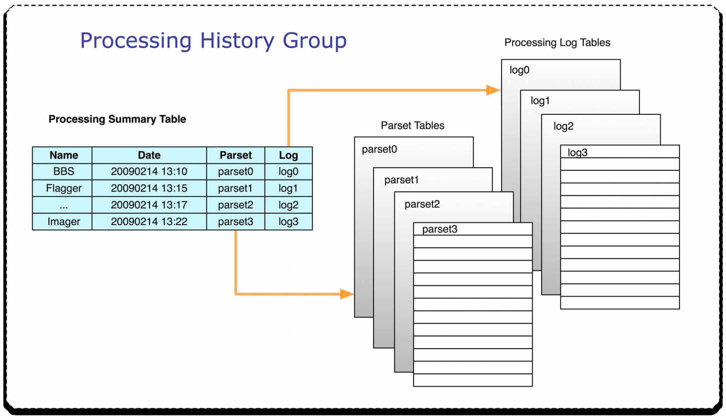

Figure 3: The processing history group, nested tabulation

As with all other RMSC file HDF5 groups and subgroups, the PROCESS_HISTORY group will be an HDF5 group, as a subgroup of an RMIMAGE 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..

5 Interfaces

—/—

5.1 Interface requirements

—/—

5.2 Relation to other workpackages

—/—

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	The source group should in addition to the LSM used contain the clean component list. Update the normal image cube too.	done
02	The peak and integrated flux densities need to be listed for all 4 Stokes parameters...or the Stokes index needs to be explicitly added.	done
03	Table 6: Add units to all descriptions, and enforce those units.	open
04	Table 6: Repeat this table. The current table is the LSM used for processing. The additional table is the clean component model found during the processing. The clean component table should be inserted into both the image cube ICD and the RM synthesis cube ICD.	na

A.2 Future enhancements

The current version of this ICD has been designed to provide the user with a basic set of attributes and metadata for describing a RMSC data product. In general, a complete description will require additional information. A discussion of features which are currently in development follows:

- **Supplemental Image Groups (SIGs)**. Supplemental Image Group will hold a variety of supplementary image products required for further analysis of rotation measure synthesis data. These SIGs are expected to contain images that otherwise cannot be derived from the data within the RMSC. For example, in most cases analysis of RM data will require some knowledge of the Stokes I information so a subset of the Stokes I image cube and a map of spectral index could be stored in an SIG. Noise/uncertainty maps are another possibility. The SIGs will not differ significantly from the RMImage group, although the coordinate type requirements will be more general (a frequency axis should be allowed for instance) and some additional attributes will likely be needed in some places.
- **Expanded PSF definition**. In general, the PSF of a RMSC is defined in 3D and in the future software will provide the PSF in such a manner. The point spread function also, generally, can be position dependent. Future implementations of the RMSC data product will provide the user with the ability to store a full 3D point spread function, and also to store several PSFs corresponding to different locations on the sky.
- **Noise Properties**. We would also like to provide for the storage of basic noise properties, derived during processing of the image cube and RMSC data.

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.

D Source list background material

- BBS supports the creation of a catalog file which a human readable text file that contains a list of sources:

```
3CR2,     POINT,  00:03:49,    -00.23.00,     15.0,  0.0,  0.0,  0.0
3CR6.1,   POINT,  00:14:12,    +79.24.00,     16.0,  0.0,  0.0,  0.0
3CR9,     POINT,  00:17:50,    +15.23.00,     15.0,  0.0,  0.0,  0.0
3CR10,    POINT,  00:22:48.1,  +63.50.42.0,  134.0,  0.0,  0.0,  0.0
3CR11.1,  POINT,  00:27:06,    +63.24.00,     11.5,  0.0,  0.0,  0.0
3CR13,    POINT,  00:31:35,    +39.03.00,     10.5,  0.0,  0.0,  0.0
```

  The fields supported by BBS are:

  - `Name`
  - `Type` (source type, either `POINT` or `GAUSSIAN`)
  - `Ra`, `Dec` (position)
  - `I`, `Q`, `U`, `V` (Stokes parameters)
  - `ReferenceFrequency` (frequency for which I, Q, U, V are given)
  - `SpectralIndexDegree` (degree of the spectral index approximation, 0 for constant spectral index)
  - `SpectralIndex:0..N` (coefficients for the spectral index approximation)
  - `MajorAxis`, `MinorAxis` (FWHM for a Gaussian source in arcsec)
  - `Orientation` (orientation of the major axis in degrees North over East)

  The fields `Name`, `Type`, `Ra`, and `Dec` are mandatory. The only source model types currently supported are `POINT` (point source) and `GAUSSIAN` (Gaussian source). The fields `MajorAxis`, `MinorAxis`, and `Orientation` are only meaningful for the Gaussian source model. For Gaussian sources, `I` is assumed to be the total flux.

- The table below is taken from the Byte-by-byte description of the 3C catalog of radio sources and serves as orientation for the construction of the source table.

Bytes	Format	Units	Label	Explanations
1—5	A5	—	`3CR`	3CR source designation (1)
8—9	I2	h	`RAh`	Right ascension (1950.0)
11—12	I2	min	`RAm`	Right ascension (1950.0)
14—17	F4.1	s	`RAs`	Right ascension (1950.0)
20—23	F4.1	s	`e_RAs`	rms uncertainty on right ascension
25	A1	—	`DE-`	Declination sign (1950.0)
26—27	I2	deg	`DEd`	Declination (1950.0)
29—32	F4.1	arcmin	`DEm`	Declination (1950.0)
34—39	F6.2	arcmin	`e_DEm`	rms uncertainty on declination
41—43	I3	deg	`GLON`	Galactic longitude
46—48	I3	deg	`GLAT`	Galactic latitude
50—56	F7.1	Jy	`S178MHz`	Flux density at 178MHz
57	A1	—	`n_S178MHz`	[*] 469λ interferometer (2)
61	A1	—	`l_Diam`	Limit flag on Diam
62—65	F4.1	—	`Diam`	? Angular diameter
66	A1	—	`x_Diam`	["'] Units of diameter (' or ")
67—70	F4.1	—	`e_Diam`	? rms uncertainty on Diam
71	A1	—	`f_Diam`	[)] Confusing sources
73—208	A136	—	`Notes`	Notes

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