LOFAR Data Format ICD
Dynamic Spectrum Data

Document ID: LOFAR-USG-ICD-006

Version 2.03.04

SVN Repository Revision: 8550

J.-M. Grießmeier, A. Alexov, K. Anderson, L. Bähren

SVN Date: 2011-10-25

Contents

Change record

Version	Date	Sections	Description of changes
0.1	2009-09-28	all	document creation.
0.2	2009-11-02	all	update file names and format (dynspec).
0.3	2009-11-18	all	adjusted document to standard structure.
0.4	2010-02-09	all	rewrote document (not complete).
0.5	2010-02-12	all	continued rewriting of document.
0.6	2010-02-19	all	first complete version.
0.65	2010-03-08	all	minor updates, different keywords. Repl. fig. 1, 2.
0.70	2010-04-19	4	Refactor section 4.1 == 4.1.1, 4.1.2
0.8	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").
0.9	2010-07-02	all	many minor modification and corrections.
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.
2.00.02	2010-10-28	all	minor modification and corrections; change of naming scheme to "sub-array pointing" and "beam".
2.00.03	2010-11-23	all	replaced "unit" by "value" where appropriate.
2.01.00	2011-03-04	all	Fixed nomenclature of ATTRIBUTES, GROUPS and DATASETS. Expanded glossary.
2.01.01	2011-03-08	all	Changed Fig. 4 & 3. Changed DynSpec to DYN_SPEC.
2.01.02	2011-03-10	all	Maintain list of references through BibLaTeX database.
2.02.00	2011-04-19	4.5	renamed "Frequency" to "Spectral". Reworked section 4.5 on coordinates.
2.03.00	2011-04-27	??	Attributes describing shape of data array; fixing section name.
2.03.01	2011-05-11	??	Added paragraph with notation conventions.
2.03.02	2011-05-23	all	small changes in notation; moved section 6 to appendix; inverted sections 3.2 and 3.3
2.03.03	2011-06-29	all	Matching up group type attributes and notation.
2.03.04	2011-07-06	all	Matching up group type attributes and notation; consolidation of labels to refer to standard sections and tables.
2.03.05	2011-10-25	all	Changed all 'float' types to 'double'; changed all 'bool' types to 'unsigned int'; changed all 'integer' to 'int'.

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

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

where

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

Notation.

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

Acknowledgements

We would like to thank all those who contributed to this document by proof-reading, providing valuable suggestions, corrections, comments, and advice. In particular, we would like to thank Frank Breitling and Jochen Eislöffel.

1 Introduction

1.1 Purpose and Scope

This interface control document (ICD) describes the data format for LOFAR dynamic spectrum data. It was derived from the ICDs describing Beam-Formed Data [?] and Sky Images [?].

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 Dynamic Spectrum Data file will be the data hosting structure for dynamic spectrum data produced by LOFAR, irrespective of their scientific purpose. It is one of the tasks of the LOFAR project to define and describe the structure of the LOFAR Dynamic Spectrum Data file format.

Dynamic spectra will be useful e.g. for planetary data, flare stars, the Sun, terrestrial lightning, and pulsars. The typical application is to have one or several beams (e.g. 'ON' and 'OFF' beams) within one sub-array pointing. Each of these beams produces a dataset accompanied by a number of by-products, such as flagging information, event tables, 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 Dynamic Spectrum Data 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 also for the associated and related meta-data describing the file 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 Dynamic Spectrum Data 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.
`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

LOFAR 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. Dynamic Spectra, 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 data array dimensions that LOFAR 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 "Dynamic Spectrum". The dataset array in a Dynamic Spectrum will be a `ndarray` data structure, as can be created by the C-based *numarray/numpy* Python packages. The Dynamic Spectrum is designed to store the polarized intensity data produced from the Dynamic Spectrum pipeline. The nominal dimensionality of a Dynamic Spectrum `Data` group's dataset will be 3 (`N_AXIS=3`), wherein the Dynamic Spectrum cube (or cubes) will be defined in (C-type order) *Polarization*, *time*, *spectral*, 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 Dynamic Spectrum Data files, and the various group entities and sub-structures comprising these files, i.e. LOFAR group types, units, physical quantities.

3 Organization of the data

3.1 High level LOFAR Dynamic Spectrum Data file structure

A LOFAR Dynamic Spectrum Data file will adhere to the following guidelines:

A LOFAR Dynamic Spectrum Data 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 Dynamic Spectrum Data file is designed to be a "flat" as possible, providing access to the necessary data without undue hierarchical tree crawling.

Therefore, the Dynamic Spectrum Data file 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.

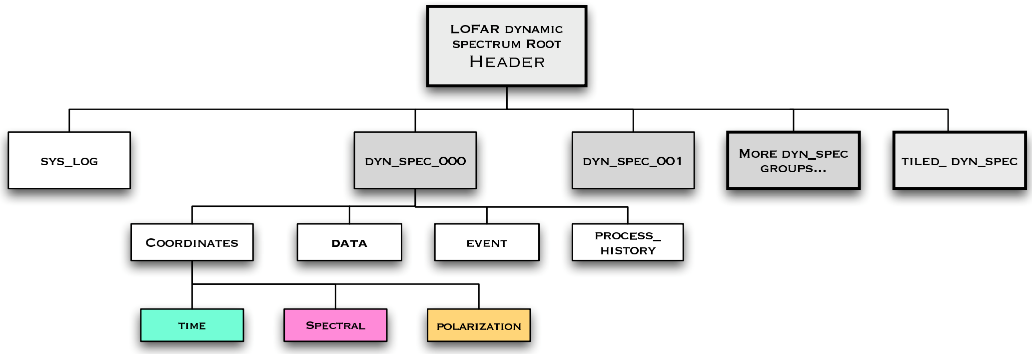

Figure 1: Dynamic Spectrum Data file structure.

3.2 Overview of Dynamic Spectrum Groups

The layout of a LOFAR Dynamic Spectrum Data file is shown in Figure ??. A LOFAR Dynamic Spectrum Data file will comprise a `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 Dynamic Spectrum Data file will contain an arbitrary, observation-dependent number of `DYN_SPEC Groups` containing a `COORDINATES Group`, an `EVENT Group`, and a `PROCESS_HISTORY Group`, which contains pertinent logs and parameter sets of the relevent image sub-band.

The main building blocks of the Dynamic Spectrum Data HDF5 file are:

1. **File Root-level** (`ROOT`). The ROOT level of the file contains the majority of associated meta-data, describing the circumstances of the observation. These data attributes include observation time (start and end), frequency window (high band vs. low band, filters) and other important characteristics of the dataset. See Sections **??** and **??** for details.
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 details.
3. **DYN_SPEC Groups**. Each observation dynamic spectrum is stored as a separate group within the file, containing its own set of four sub-groups. Characteristics about each dynamic spectrum are stored as Attributes in group headers. A LOFAR Dynamic Spectrum Data file may contain numerous dynamic spectra groups. Possibilites include:
   - one dynamic spectrum containing the data from all stations where different stations all observe at the same frequencies
   - one dynamic spectrum per station when different stations observe at different frequencies (in that case, the combined spectrum may be written to the `TILED_DYN_SPEC Group`, see below),
   - one dynamic spectrum per station where different stations observe at the same frequencies (e.g. to monitor RFI),
   - separate dynamic spectra for ON and OFF beams,
   - dynamic spectra created from the same file, but with different frequency and time resolution.

   Each `DYN_SPEC Group` will contain it's own `COORDINATES Group` (see below) plus one `Data` group, which will in turn contain a dataset as an ndarray, along with associated attributes. See Section **??** for details.
4. **EVENT Groups**. In some (if not all) cases, dynamic spectra will be monitored for events (flares, bursty emission, ...) during file creation. In those cases, a **EVENT Group** will contain a table of events. See Section **??** for details.

5. **Coordinates Groups** (COORDINATES). Each DYN_SPEC Group contains one COORDINATES Group, which contain 3 Coordinate sub-groups (the TIME, SPECTRAL and POLARIZATION coordinates). See Section **??** for details.
6. **Processing History Groups** (PROCESS_HISTORY) can be found on the DYN_SPEC Group level. These are catch-all envelops encapsulating information about all the steps of processing, such as parameter sets and processing logs. See Section **??** for details.
7. **Dynamic Spectrum DATA arrays**. For each DYN_SPEC Group, the dynamic spectra 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. See Section **??** for details.
8. **TILED_DYN_SPEC Group**. In the case where different stations observe at different frequencies, a TILED_DYN_SPEC Group can be used to accomodate a composite dynamic spectrum, combining the frequency information of different DYN_SPEC groups. See Section **??** for details.

3.3 Hierarchical Structure of the HDF5 file

The Dynamic Spectrum Data are organized within a hierarchical structure, which reflects upon the structure in which data are grouped during processing. This structure can be represented as an HDF5 hierarchy in the following way:

```
OBS_NUMBER/
OBS_NUMBER/SYS_LOG
OBS_NUMBER/DYN_SPEC_000/COORDINATES
OBS_NUMBER/DYN_SPEC_000/COORDINATES/TIME_COORD
OBS_NUMBER/DYN_SPEC_000/COORDINATES/SPECTRAL_COORD
OBS_NUMBER/DYN_SPEC_000/COORDINATES/POLARIZATION_COORD
OBS_NUMBER/DYN_SPEC_000/DATA
OBS_NUMBER/DYN_SPEC_000/EVENT
OBS_NUMBER/DYN_SPEC_000/PROCESS_HISTORY
...
OBS_NUMBER/DYN_SPEC_NNN/...
OBS_NUMBER/TILED_DYN_SPEC/...
...
```

An alternate way to display the above described HDF5 hierarchy is:

```
 OBS_NUMBER
|-- SYS_LOG
|-- DYN_SPEC_000
|   |-- COORDINATES
|   |   |-- TIME_COORD
|   |   |-- SPECTRAL_COORD
|   |   `-- POLARIZATION_COORD
|   |-- DATA
|   |-- EVENT
|   `-- PROCESS_HISTORY
|
|-- DYN_SPEC_NNN
|   |
|   ...
|
`-- TILED_DYN_SPEC
    |
    ...
```

3.4 Dynamic Spectrum Data flow

The dynamic spectrum data flow is shown in Figure **??**.

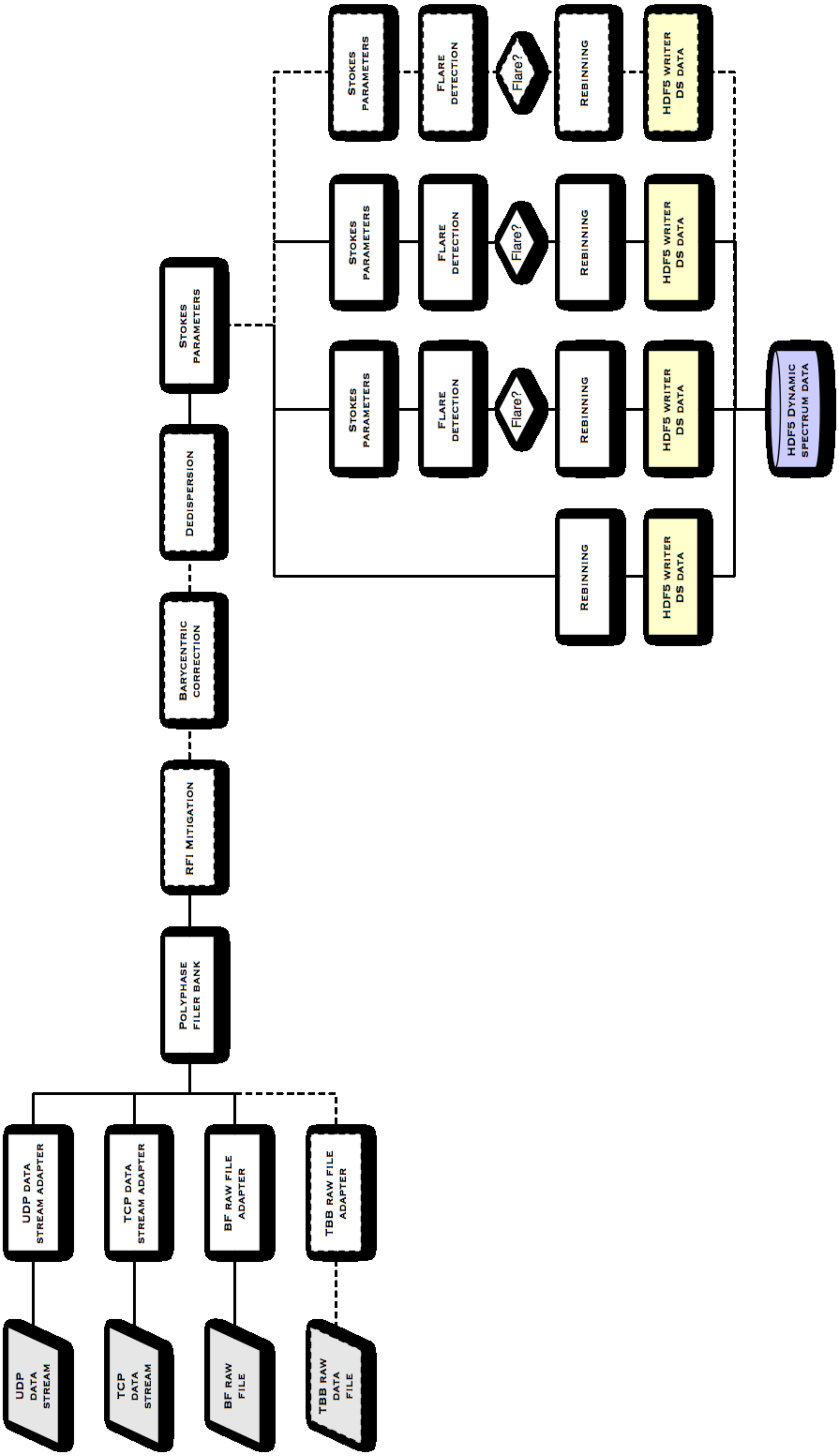

Figure 2: Dynamic Spectrum Data Flow

4 Detailed Data Specification

4.1 The ROOT Group

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 Dynamic Spectrum 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 dynamic spectra. 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 dynamic spectra-specific.

In other words,

```
ROOT Attributes =
Common LOFAR Attributes (CLA) + Additional (dynamic spectrum specific) ROOT Attributes.
```

The Common LOFAR Attributes are the first attributes of any LOFAR **File Root Group**.

Table **??** lists the group types used for LOFAR Dynamic Spectrum Data.

4.1.1 Common LOFAR Attributes

This section will specify a set of ROOT-level attributes that will be common to all LOFAR science data products. These "LOFAR common metadata" will appear as attributes at the ROOT level of all LOFAR Dynamic Spectrum Data files, as well as *all* other LOFAR products. These common LOFAR metadata 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 LOFAR Dynamic Spectrum Data with the file's 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.

For LOFAR Dynamic Spectrum Data, FILETYPE='dyn_spec'.

- 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,

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.

  - 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:

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

  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.

- `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	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.

- ANTENNA_SET – The **antenna set** configuration used during the observation; see Table **??** below for a list of recognized values.
- 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	`'SYS_LOG'`	System log files, parsets
TILED_DYN_SPEC	`'TILED_DYN_SPEC'`	Tiled Dynamic Spectrum
DYN_SPEC	`'DYN_SPEC'`	Dynamic spectrum
DYN_SPEC Subgroups	**Value**	**Description**
Data group	`'DATA'`	This is a Data group of a Dynamic Spectrum Data file
Event group	`'EVENT'`	This is a Event List group
Processing History group	`'PROCESS_HISTORY'`	This is a Processing History group
Coordinates Group	`'COORDINATES'`	This is a Coordinates group
Coordinates Group Subgroups	**Value**	**Description**
Time coord group	`'TIME_COORD'`	This is a coord group
Spectral coord group	`'SPECTRAL_COORD'`	This is a coord group
Polarization coord group	`'POLARIZATION_COORD'`	This is a Stokes coordinate group

Table 7: LOFAR Dynamic Spectrum Data Group Types.

4.1.2 Additional dynamic spectrum ROOT Attributes

Table **??** contains the Dynamic Spectrum Data ROOT header Attributes. The CLA Attributes have already been listed in Section **??** above, therefore these are additional attributes in the ROOT header.

FIELD/KEYWORD	H5TYPE	TYPE	UNIT	VALUE	DESCRIPTION
`DYN_SPEC_GROUPS`		unsigned int	—	`'true'`	File has `DYNC_SPEC` subgroups
`NOF_DYN_SPEC`		int	—	—	number of `DYNC_SPEC groups` in this file
`CREATE_OFFLINE_ONLINE`	Attr.	`string`	—	—	Whether the file was created 'Online' (stream to file) or 'Offline' (file to file).
`BF_FORMAT`	Attr.	`string`	—	—	"RAW" if the file is BeamFormed RAW data; "TAB" if the file is BeamFormed Processed data (will usually be "TAB").
`BF_VERSION`	Attr.	`string`	—	—	BeamFormed data format version number.
`NOF_STATIONS`	Attr.	`int`	—	—	Number of stations used within this Beam
`STATIONS_LIST`	Attr.	`array<string,1>`	—	—	List of stations used for this Beam; list must match the NOF_STATIONS numerically
`PRIMARY_POINTING_DIAMETER`	Attr.	`double`	arcmin	—	FWHM of the sub-array pointing at zenith at center frequency.
`POINT_RA`	Attr.	`double`	deg	—	J2000 right ascension of sub-array pointing at start of observation, in degrees (at LOFAR core).

continued on next page

Field/Keyword	H5Type	Type	Unit	Value	Description
POINT_DEC	Attr.	double	deg	—	J2000 declination of station beam at start of observation, in degrees (at LOFAR core).
POINT_ALTITUDE	Attr.	array<string,1>	deg	—	Altitude of the pointing at start of observation, per stations list, in the same order.
POINT_AZIMUTH	Attr.	array<string,1>	deg	—	Azimuth of the pointing at start of observation, per stations list, in the same order.
CLOCK_RATE	Attr.	double	—	—	Clock rate, in units of CLOCK_RATE_UNIT (MHz)
CLOCK_RATE_UNIT	Attr.	string	'MHz'	—	Clock rate units
NOF_SAMPLES	Attr.	int	—	—	number of time samples.
SAMPLING_RATE	Attr.	double	—	—	Sampling rate, in units of SAMPLING_RATE_UNIT
SAMPLING_RATE_UNIT	Attr.	string	'MHz'	—	Sampling rate units
SAMPLING_TIME	Attr.	double	—	—	Sampling time is 1/SAMPLING_RATE, in units of SAMPLING_TIME_UNIT
SAMPLING_TIME_UNIT	Attr.	string	'μs'	—	Sampling time units
TOTAL_INTEGRATION_TIME	Attr.	double	—	—	Total integration time = SAMPLING_TIME * NOF_SAMPLES
TOTAL_INTEGRATION_ TIME_UNIT	Attr.	string	's'	—	Total integration time units
CHANNELS_PER_SUBBAND	Attr.	int	—	—	Number of channels for each subband
SUBBAND_WIDTH	Attr.	double	—	—	Subband width 0.156250 or 0.1953125
SUBBAND_WIDTH_UNIT	Attr.	string	'MHz'	—	Subband width units
CHANNEL_WIDTH	Attr.	double	—	—	Channel width is equal to the SUBBAND_WIDTH / CHANNELS_PER_SUBBAND
CHANNEL_WIDTH_UNIT	Attr.	string	'MHz'	—	Channel width units
TOTAL_BANDWIDTH	Attr.	double	MHz	—	Total bandwidth (excluding gaps).
WEATHER_STATIONS_LIST	Attr.	array<string,1>	—	—	List of stations with weather information.
WEATHER_TEMPERATURE	Attr.	array<double,1>	$^{\circ}C$	—	Approximate outside temperature; order must match the listing in WEATHER_STATIONS_LIST attribute. (* See note below)
WEATHER_HUMIDITY	Attr.	array<double,1>	—	—	Approximate humidity (%); order must match the listing in WEATHER_STATIONS_LIST attribute (* See note below)

continued on next page

FIELD/KEYWORD	H5TYPE	TYPE	UNIT	VALUE	DESCRIPTION
SYSTEM_TEMPERATURE	Attr.	array<double,1>	K	—	System temperature for the various stations ; order must match the listing in WEATHER_STATIONS_LIST attribute. (* See note below)

Table 8: Additional ROOT group attributes for LOFAR Dynamic Spectrum Data.

**WEATHER_TEMPERATURE, WEATHER_HUMIDITY and SYSTEM_TEMP are not available for all stations.*

4.2 The System Logs Group (SYS_LOG)

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

4.3 The DYN_SPEC Group

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

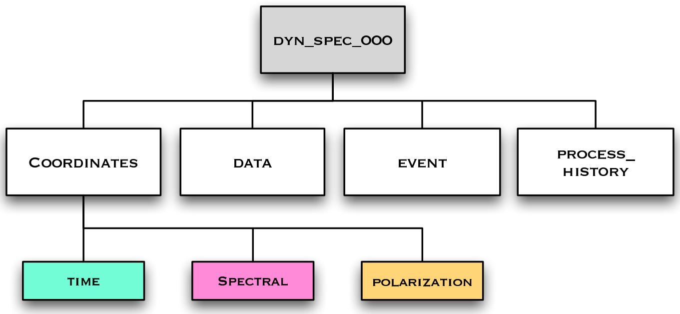

Figure 3: DYN_SPEC group structure.

A DYN_SPEC group *(Fig. 2)* will comprise four sub groups. These four groups, two hierarchical levels below the "ROOT group" in a LOFAR Dynamic Spectrum Data file *(Fig. 1)*, will be

- A COORDINATES group that will contain one subgroup of each of the following kinds: TIME_COORD, SPECTRAL_COORD, POLARIZATION_COORD, which describe various axes of the associated dataset.
- A DATA group that will contain a dataset array.
- An EVENT group that will be a tabular list of localized events (flares, bursts, etc.).
- 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..

Figure 2 illustrates the form of an `DYN_SPEC Group` in a LOFAR Dynamic Spectrum Data file. Most of the relevent `DYN_SPEC Group` metadata will be contained within the Coordinates groups (*see* **??**, *"Coordinates group"*)

Field/Keyword	H5Type	Type	Unit	Description
GROUPTYPE	Attr.	string	'DYN_SPEC'	LOFAR group type
DYN_SPEC_START_MJD	Attr.	string	NNNNNN.NNNNNNN	Start time of obs (MJD).
DYN_SPEC_STOP_MJD	Attr.	string	NNNNNN.NNNNNNN	End time of obs (MJD).
DYN_SPEC_START_UTC	Attr.	string	hh:mm:ss.sssssssss	Start time of obs (UTC time, 24-h clock).
DYN_SPEC_STOP_UTC	Attr.	string	hh:mm:ss.sssssssss	End time of obs (UTC time, 24-h clock).
DYN_SPEC_START_TAI	Attr.	string	hh:mm:ss.sssssssss	Start time of obs (International Atomic Time).
DYN_SPEC_STOP_TAI	Attr.	string	hh:mm:ss.sssssssss	End time of obs (International Atomic Time).
DYN_SPEC_BANDWIDTH	Attr.	double	MHz	Bandwidth (excluding gaps).
BEAM_DIAMETER	Attr.	double	arcmin	FWHM of the beams at zenith at center frequency.
TRACKING	Attr.	string	–	'J2000' = tracking ON; 'LMN' = tracking OFF; TBD = tracking of solar system object
TARGET	Attr.	array<string,1>	—	targets/sources observed within this Beam
ONOFF	Attr.	string	—	e.g. 'ON' or 'OFF_0', 'OFF_1', ...
POINT_RA	Attr.	double	—	J2000 right ascension of the center of the beam at start of observation, in degrees (at LOFAR core).
POINT_DEC	Attr.	double	—	J2000 declination of the center of the beam at start of observation, in degrees (at LOFAR core).
POSITION_OFFSET_RA	Attr.	double	—	RA offset of beam from sub-array pointing at start of observation (degrees).
POSITION_OFFSET_DEC	Attr.	double	—	Dec offset of beam from sub-array pointing at start of observation (degrees).
BEAM_DIAMETER_RA	Attr.	double	—	The diameter of the beam ellipse in the major axis a (RA), units of degrees.
BEAM_DIAMETER_DEC	Attr.	double	—	The diameter of the beam ellipse in the minor axis b (Dec), units of degrees.
BEAM_FREQUENCY_MAX	Attr.	double	—	Beam maximum frequency
BEAM_FREQUENCY_MIN	Attr.	double	—	Beam minimum frequency
BEAM_FREQUENCY_CENTER	Attr.	double	—	Beam center frequency: $0.5 \cdot (\max + \min)$.
BEAM_FREQUENCY_UNIT	Attr.	string	'MHz'	Units for BEAM_FREQUENCY_MAX, BEAM_FREQUENCY_MIN and BEAM_FREQUENCY_CENTER.

continued on next page

Field/Keyword	H5Type	Type	Unit	Description
BEAM_NOF_STATIONS		int	—	N. of stations used for this beam
BEAM_STATIONS_LIST		array<string,1>	—	List of stations used for this beam
DEDISPERSION	Attr.	string	—	Were the data dedispersed incoherently (INCOHERENT), dedispersed coherently (COHERENT), or (usually) not dedispersed (NONE)?
DISPERSION_MEASURE	Attr.	double	—	The dispersion measure applied to the data, if data were dedispersed
DISPERSION_MEASURE_UNIT	Attr.	string	‘pc/cm^3’	The dispersion measure units.
BARYCENTER	Attr.	unsigned int	—	Are the data barycentered? (True=1=Yes/False=0=No). Default is 0.
STOKES_COMPONENTS	Attr.	array<string,1>	—	Stokes components for which data are attached to this group.
COMPLEX_VOLTAGE	Attr.	unsigned int	—	Is the data in Complex Voltages (Xreal Ximg, Yreal, Yimg)? (True=1=Yes/False=0=No)
SIGNAL_SUM	Attr.	string	—	was the signal summed coherently (COHERENT) or incoherently (INCOHERENT_SKY, INCOHERENT_PRIMARY_POINTING)?

Table 9: DYN_SPEC Group Attributes.

Some of these keywords need further clarification:

- POINT_RA & POINT_DEC – J2000 right ascension and declination of the center of the beam at start of observation, in degrees (at the LOFAR core, as the positions will be slightly different for solar system objects).
- STOKES_COMPONENTS – Providing both flexibility and readability, the Stokes components for which data are attached to the Beam Group, are stored as array<string,1>. The attribute has the flexibility to host different combinations and types of Stokes data, such as e.g.

  STOKES_COMPONENTS = ["I"]
  STOKES_COMPONENTS = ["I","Q","U","V"]
  STOKES_COMPONENTS = ["I","Q"]
  STOKES_COMPONENTS = ["L","R"]
  STOKES_COMPONENTS = ["XX","XY","YX","YY"].

4.4 The TILED_DYNSPEC Group

This group has the same structure as the DYN_SPEC groups. In the case where different stations observe at different frequencies, this group can be used to accomodate a composite dynamic spectrum, combining the frequency information of different DYN_SPEC groups.

> **Comment:**
> The final format is not yet defined. This group should include information from which DynSpec groups the information as taken.

4.5 The Coordinates Group (COORDINATES)

The Coordinates Group acts as a container to take up a collection of coordinates, as described in the subsequent sections below. Besides this function as a container – grouping together embedded coordinate objects – the Coordinates Group also provides basic reference frame information, which is required for the proper transformation of quantities to other reference systems.

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, as presented in Table ??, summarize the overall characteristics of the set of coordinates collected within a COORDINATES Group. Specifially, for a LOFAR DYN_SPEC Group, the following values are used:

- NOF_COORDINATES – The number of coordinate objects/groups contained within the coordinates group. For Dynamic Spectrum Data, this is 3 (TIME_COORD, SPECTRAL_COORD, POLARIZATION_COORD).
- NOF_AXES – The number of coordinate axes associated with the coordinate objects. For Dynamic Spectrum Data, the number of axes is will be 3, i.e. 1 for each of the subgroups TIME_COORD, SPECTRAL_COORD, POLARIZATION_COORD.

The different axes stored within this container are described in the remainder of this section. More information on coordinates in LOFAR data files in general can be found in ICD-002 [?].

4.5.1 Time coordinate

This group describes the time coordinate of the associated DYN_SPEC Group.
Specifially, for a LOFAR DYN_SPEC Group, the following values are used:

- COORDINATE_TYPE — The coordinate type = 'Time'.
- STORAGE_TYPE — The storage type can be either 'Linear' or 'Tabular'. For dynamic spectra, it will be 'Linear' in most cases.
- NOF_AXES — The number of coordinate axes = 1.
- AXIS_NAMES — The names of the axis = 'Time'.
- AXIS_UNITS — The units of the coordinate axis, e.g. = 'microseconds'.
- REFERENCE_VALUE — The World Coordinate Reference value (CRVAL) will be the 0,0 location of the dataset (start time of each sample).
- REFERENCE_PIXEL — The World Coordinate Reference pixel (CRPIX) will be the 0,0 location of the dataset.

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

Table 11: Recognized values for the reference frame to specify a location; values and descriptions have been adopted from the "Space-Time Coordinate Metadata for the Virtual Observatory" [?], as produced by the IVOA Data Model Working Group.

- INCREMENT — The World Coordinate increment (CDELT) is the time bin (SAMPLING_TIME).
- PC — The World Coordinate Reference scaling delta matrix is flat (1, 0) for Dynamic Spectrum data.

More detail on time coordinates for LOFAR data files can be found in Section 4.3.1 of ICD-002 [?].

4.5.2 Spectral coordinate

This group describes the spectral coordinate of the associated DYN_SPEC Group.

Specifially, for a LOFAR DYN_SPEC Group, the following values are used:

- COORDINATE_TYPE — The coordinate type = 'Spectral'.
- STORAGE_TYPE — The storage type can be either 'Linear' or 'Tabular'. For dynamic spectra, it will be 'Tabular' in most cases[1].
- NOF_AXES — The number of coordinate axes = 1.
- AXIS_NAMES — The names of the axis = 'Frequency'.
- AXIS_UNITS — The units of the coordinate axis is = 'MHz'.
- REFERENCE_VALUE — The World Coordinate Reference value (CRVAL) will be the 0,0 location of the dataset.
- REFERENCE_PIXEL — The World Coordinate Reference pixel (CRPIX) will be the 0,0 location of the dataset.

[1] Given the flexibility concerning the arrangement of frequency channels or subbands, the values along this coordinate axis might be linear, but do not necessarily have to be.

Field/Keyword	H5Type	Type	Value	Description
`GROUPTYPE`	Attr.	`string`	`'TimeCoord'`	Group Type descriptor
`COORDINATE_TYPE`	Attr.	`string`	`'Time'`	Coordinate Type descriptor
`STORAGE_TYPE`	Attr.	`array<string,1>`	`'Linear'` | `'Tabular'`	coordinate storage type
`NOF_AXES`	Attr.	`int`	1	N of coordinate axes
`AXIS_NAMES`	Attr.	`array<string,1>`	[‘Time’]	World axis names
`AXIS_UNITS`	Attr.	`array<string,1>`	[‘s’]	World axis units
`REFERENCE_VALUE`	Attr.	`array<double,1>`	—	Reference value
`REFERENCE_PIXEL`	Attr.	`array<double,1>`	—	Reference pixel
`INCREMENT`	Attr.	`array<double,1>`	—	Coordinate increment
`PC`	Attr.	`array<double,1>`	—	
`AXIS_VALUES_PIXEL`	Attr.	`array<double,1>`	—	Reference pixels
`AXIS_VALUES_WORLD`	Attr.	`array<double,1>`	—	Reference values

Table 12: Time Coordinate Attributes

Field/Keyword	H5Type	Type	Value	Description
`GROUPTYPE`	Attr.	`string`	‘SpectralCoord’	Group type descriptor
`COORDINATE_TYPE`	Attr.	`string`	‘Spectral’	Coordinate Type descriptor
`STORAGE_TYPE`	Attr.	`array<string,1>`	`'Linear'` | `'Tabular'`	coordinate storage type
`NOF_AXES`	Attr.	`int`	1	nof. coordinate axes
`AXIS_NAMES`	Attr.	`array<string,1>`	[‘Frequency’]	World axis names
`AXIS_UNITS`	Attr.	`array<string,1>`	[‘Hz’]	World axis units
`REFERENCE_VALUE`	Attr.	`array<double,1>`	—	Reference value (CRVAL)
`REFERENCE_PIXEL`	Attr.	`array<double,1>`	—	Reference pixel (CRPIX)
`INCREMENT`	Attr.	`array<double,1>`	—	Coordinate increment (CDELT)
`PC`	Attr.	`array<double,1>`	—	
`AXIS_VALUES_PIXEL`	Attr.	`array<double,1>`	—	Reference pixels
`AXIS_VALUES_WORLD`	Attr.	`array<double,1>`	—	Reference values

Table 13: Spectral Coordinate Attributes

FIELD/KEYWORD	H5TYPE	TYPE	VALUE	DESCRIPTION
`GROUP_TYPE`	Attr.	`string`	`'PolarizationCoord'`	Coordinate Type descriptor
`COORDINATE_TYPE`	Attr.	`string`	`'Polarization'`	Coordinate Type descriptor
`STORAGE_TYPE`	Attr.	`array<string,1>`	`'Tabular'`	coordinate storage type
`NOF_AXES`	Attr.	`int`	1	N of coordinate axes
`AXIS_NAMES`	Attr.	`array<string,1>`	'Polarization'	World axis names
`AXIS_UNITS`	Attr.	`array<string,1>`	—	World axis units

Table 14: Stokes Groups Attributes

- `INCREMENT` — The World Coordinate increment (CDELT) is the frequency bin (CHANNEL_WIDTH).
- `PC` — The World Coordinate Reference scaling delta matrix is flat (1, 0) for Dynamic Spectrum data.
- `AXIS_VALUES_PIXEL` — Reference pixels - List of the subbands. (See Sec. ??).
- `AXIS_VALUES_WORLD` — Reference values – List of the equivalent frequencies of the list of subbands. (See Sec. ??)

More detail on spectral coordinates for LOFAR data files can be found in Section 4.3.2 of ICD-002 [?].

4.5.3 Polarization coordinate

This group describes the polarization contents of the associated `DYN_SPEC Group`. For Dynamic Spectrum data, only one axis of this type will be used, i.e. `NOF_AXES`=1.

Within each `DYN_SPEC Group`, there are either one or four Stokes Groups. If the data are summed, then there is only the Stokes I information. If the data are not summed, then there are four Stokes tables (I, Q, U, V or XX, XY, YX, YY), one per polarization. Table ?? lists the Attributes in the Stokes Groups.

Specifially, for a LOFAR `DYN_SPEC Group`, the following values are used:

- `COORDINATE_TYPE` — The coordinate type = 'PolarizationCoord'.
- `STORAGE_TYPE` is the descriptor for the underlying storage type for this coordinate, of value `'Tabular'`.
- `NOF_AXES` — The number of coordinate axes = 1.
- `AXIS_NAMES` — The names of the axis is = 'Polarization'.
- `AXIS_UNITS` — The units of the coordinate axis is = 'NONE'.

More detail on polarization coordinates for LOFAR data files can be found in Section 4.3.3 of ICD-002 [?].

4.6 The Dynamic Spectrum Dataset

A dynamic spectrum `DATA` group will most often be a subgroup of a `DYN_SPEC 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, with the adoption of a so-called "Coordinates group," which contains all the relevent information, scale and unit metadata, `Data` group attributes will be limited.

The dataset array will (usually) be a 3-D data structure. The nominal dimensionality of a `Data` group's dataset will be 3 (`N_AXIS=3`), wherein the data cube (or cubes) will be defined in (C-type order) *Polarization*, *Spectral*, *Time*.

LOFAR Dynamic Spectrum Data 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 (e.g. one dynamic

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

Table 15: Attributes attached to the dynamic spectrum dataset array.

spectrum per FITS file). See § ?? "The Coordinates group," for a detailed specification of `Data` group header attributes.

The Subband/Channel information will be stored as either 1-D or N-D tables or arrays. This is where the bulk of the data reside. The general structure is such that channels are columns and time bins are rows. If using 1-D tables/arrays, then each subband will be in its own table/array. If using N-D tables/arrays, then each subband will be a plane in a data cube. Time increments are quantized and are filled (with NaN values) for gaps; the time axis is usually linear. For each subband, the channels are quantized and filled for gaps, so within each subband table/array, the frequency axis is linear. However, since the subbands can have large gaps, the overall frequency axis is not linear. The frequency coordinate axis information is stored in the Spectral Coordinate Group in the Dynamic Spectrum Data file.

For example, Subband 0 can start at 140 MHz, Subband 1 at 150 MHz and Subband 2 at 165 MHz; each Subband has the same number of channels (subdivisions), say 512, all with the same channel widths, and same number of time increments (rows), say 2000000. Therefore, there is a gap between Subbands 1 and 2 in this example. One cannot assume that the last channel in Subband (N-1) is followed directly in frequency by the first channel in the next Subband N. Gaps are accounted for in the Spectral Coordinate Group.

There are several possibilities of sub-observing modes. Five sub-modes are listed below, but please note that this is not yet a full set of sub-modes.

1. *Raw data*

   The default data stored in each subband table is full resolution, polarized voltages from the dipoles. These voltages are stored as 16-bit complex pairs for both X and Y. In other words, each sample is stored as (X-real, X-imaginary)(Y-real, Y-imaginary) for a total of 64-bits (2 32-bit complex numbers). This is the format of the BF RAW data.

2. *Stokes I with IncoherentSum*

   ```
   STOKES_COMPONENTS = ["I"]
   SIGNAL_SUM = INCOHERENT
   ```

   Data taken for **Stokes I** TiedArray observing mode with Incoherent Summing. Data are being stored as one number, the total intensity per unit time. Currently this number is being stored as a single float and will eventually be stored as a 16-bit integer. Example HDFview TBD.

3. Stokes I with CoherentSum

   ```
   STOKES_COMPONENTS = ["I"]
   SIGNAL_SUM = COHERENT
   ```

   Data taken for **Stokes I** TiedArray observing mode with Coherent Summing. Data are being stored as one number, the total intensity per unit time. Currently this number is being stored as a single float and will eventually be stored as a 16-bit integer.

4. Full Stokes with IncoherentSum

   ```
   STOKES_COMPONENTS = ["I","Q","U","V"]
   SIGNAL_SUM = INCOHERENT
   ```

Data taken for **Full Stokes** TiedArray observing mode with Incoherent Summing. Data are being stored as 4 numbers, per unit time. Currently these 4 numbers are being stored as 4 floats and will eventually be stored as 4 16-bit integers. Example HDFview TBD.

5. Full Stokes with CoherentSum

```
STOKES_COMPONENTS = ["I","Q","U","V"]
SIGNAL_SUM = INCOHERENT
```

Data taken for **Full Stokes** TiedArray observing mode with Coherent Summing. It is stored as 4 numbers, per unit time. Currently these 4 numbers are being stored as 4 floats and will eventually be stored as 4 16-bit integers.

4.7 The EVENT Group

The `Event` group in a Dynamic Spectrum Data file will be a table of events detected within the dynamic spectrum, including their associated parameters. The `Event` group header will specify the fields (columns) of the table, and the number of events in the table (rows).

[**Comment:**
The precise format is not yet defined. This should also include information on how the events were detected, i.e. the settings used for event trigging.]

Field/Keyword	H5Type	Type	Value	Description
`GROUPTYPE`	Attr.	string	`'Event'`	Dynamic Spectrum group type
`DATASET`	Attr.	string	`'Event List'`	
`N_AXIS`	Attr.	int	2	Number of data axes
`N_AXIS_1`	Attr.	string	`'Fields'`	Axis of the data fields
`N_AXIS_2`	Attr.	string	`'Event'`	Axis of the event rows.
`N_SOURCE`	Attr.	int		Number of data rows/sources
`FIELD_1`	Attr.	double		time
`FIELD_2`	Attr.	double		frequency
`FIELD_3`	Attr.	double		Peak Flux
`FIELD_4`	Attr.	double		Integrated Flux
`FIELD_5`	Attr.	double		duration
`FIELD_6`	Attr.	double		$f_{max} - f_{min}$

Table 16: Attributes of an Event group.

4.8 The Processing History Group (PROCESS_HIST)

The data definition for the `PROCESS_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 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.

[**Comment:**
The Figure has to be adapted to DYNSPEC case.]

Field/Keyword	Type	Value	Description
GROUPTYPE	string	'PROCESS_HISTORY'	LOFAR group type
LOG_DYN_SPEC	unsigned int		Logfile from Dynamic Spectrum Writer? (True=1=Yes/False=0=No)
PARSET_OBS	unsigned int		Dynamic Spectrum parset file? (True=1=Yes/False=0=No)

Table 17: Attributes of a PROCESS_HISTORY group.

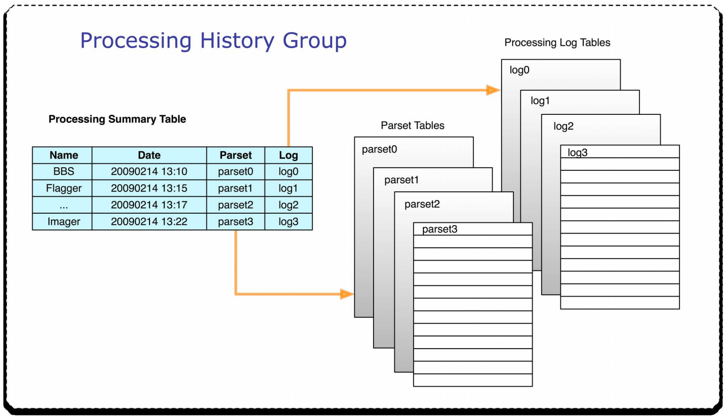

Figure 4: The PROCESS_HISTORY group, nested tabulation

As with all other Dynamic Spectrum Data file HDF5 groups and subgroups, the PROCESS_HISTORY group will be an HDF5 group, as a subgroup of a DYNAMICSPEC 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..

[**Comment:**
Include information on rebinning, etc.]

5 Interfaces

—/—

5.1 Interface requirements

—/—

5.2 Relation to other workpackages

—/—

A Discussion & open questions

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