MeasurementSet description for LOFAR
Version 2.08.01

	Organisatie / Organization	Datum / Date
Auteur(s) / Author(s): A.P. Schoenmakers, G.A.Renting	ASTRON	04-04-2018
Controle / Checked:	ASTRON	
Goedkeuring / Approval:	ASTRON	
Autorisatie / Authorisation: **Handtekening / Signature**	ASTRON	

NETHERLANDS FOUNDATION FOR RESEARCH IN ASTRONOMY

Distribution list:

Group:	Others:

Document history:

Revision	Date	Chapter / Page	Modification / Change
0.1	March 2006	-	Creation
0.2	20 July 2006		Added columns to MAIN; filled in station name descriptions in ANTENNA; transfer to current document template, filling in more details, removal of irrelevant table entries.
0.3	30 November 2006		Added MS schema figure, Added appendix on how to translate parset file to MS fields
0.4	24 January 2007		Small additions due to recent developments
0.5	25 February 2008		Updated to latest field values/formats/keywords used
0.6 *Drafts*	February-September 2010		Draft updates for better dipole and antenna field support and full LOFAR operations
2.06.07	14 October 2010		New numbering scheme and naming conventions compatible with ICDs Final version of updates for dipole and antenna field support for LOFAR operational phase
2.07.01	19 April 2011		New version with new LOFAR_ANTENNA_FIELD and LOFAR_ELEMENT_FAILURE tables and some smaller updates
2.07.02	15 September 2011		Corrected two fields that were omitted by accident in FIELD and LOFAR_ELEMENT_FAILURE tables
2.08.00	14 February 2012		Various minor corrections and addition of optional QUALITY_* tables
2.08.01	4 April 2018		Minor clarification on how polarisations are stored.

**LOFAR/
USG/
Data Formats**

Doc.nr.:

Rev.:

Date:

Class.:   Public

2 of 32

Table of contents:

	LOFAR/ USG/ Data Formats	**Doc.nr.:**
		Rev.:
		Date:
		Class.: Public

	LOFAR/	**Doc.nr.:**	
LOFAR	**USG/**	**Rev.:**	
	Data Formats	**Date:**	
		Class.:	Public

1 Introduction

In this document I present detailed schemes of the tables in a Measurement Set, and elaborate on their exact definitions. This will grow into the blueprint of a new LOFAR MS.

This is based (and partially copied) from the original AIPS++ MS 2.0 definition document (AIPS++ Note 229) and should be compatible with it. The Measurement Set v2 is described in great detail in an AIPS++ memo, which can be found at these locations:

http://aips2.nrao.edu/stable/docs/notes/229/229.html or
http://www.astron.nl/aips++/docs/notes/229/229.html or
http://www.astron.nl/casacore/trunk/casacore/doc/notes/229.html

The latest version of this document is available at:

http://www.lofar.org/wiki/doku.php?id=public:documents:lofar_documents

Some fields and tables that are not relevant for LOFAR at all have been skipped (see the MS2 definition document for their meaning).

1.1 Antenna field clarification

Please observe that LOFAR does not have traditional antennas, but instead uses fields of dipoles. These get beam formed in groups to behave as a more traditional antenna and are called Antenna Fields in this document. A station has several Antenna Fields, some of which can be active at the same time during an observation.

There is one special mode called HBA_JOINED in which basically joins two different groups of dipoles into one Antenna Field, where normally each group is a separate antenna field. This mode has required the creation of a separate LOFAR_ANTENNA_FIELD table and has made it a bit unclear what an Antenna Field exactly means. In this document I make the distinction through capitalisation: An Antenna Field corresponds to a traditional Antenna, and is the end point of a baseline. An antenna field is one of the groups of dipoles that alone or together with another antenna field form the Antenna Field. Each antenna field can be found in the LOFAR_ANTENNA_FIELD table.

1.2 Other important conventions

Unless otherwise specified, direction coordinates are specified in reference to the LOFAR Core. This is relevant because of the size of LOFAR a lot of objects like the Moon and planets are in near field.

Also take note that SubArray Pointing was referred to as Station Beam or Beam in older versions of this and other LOFAR documents. The term Beam is now reserved for what used to be called Pencil Beam. Columns that are optional in a table have their name marked in *italics* in the description.

In the format description, either a single value or an array can be specified. If an array is specified, if will be noted in round parenthesis, for example, Double(3,2) is an 3x2 array of Double. A star (*) symbol is used for a variable length array, for example String(*) is a variable length array of Strings. Please also note that the various types like Double, Float, String, Int, Complex, are well defined within the CASA/AIPS++ project in endianness and IEEE compliance. See the casa::Conversion Class Reference and aipstype.h.

Some columns in the MeasurementSet require coordinate and unit specification. This is done in a manner compatible with the CASA/AIPS++ Measures system. Measure frame information is implicit in the underlying MS data. See also the *Guide to AIPS++ Measures*, AIPS++ Note 223
http://www.astron.nl/casacore/trunk/casacore/doc/notes/233.html

All indices for antennas, feeds, spectral windows, or related quantities are assumed to start at zero.

The MeasurementSet uses the CASA/AIPS++ table system, which has a powerful query language called TAQL for accessing and manipulating the data. See *Table Query Language*, AIPS++ Note 199 for details:
http://www.astron.nl/casacore/trunk/casacore/doc/notes/199.html

2 General description of the Measurement Set

This section briefly describes all tables and sub-tables in a Measurement set, and what type of information these contain. I will discuss individual fields in so far as these require additional explanations to understand their exact meaning. These are mostly administrative fields, not the main data fields.

2.1 Structure of the MS

In the table below a short summary of the sub-tables in a LOFAR Measurement Set is presented and a brief explanation of their contents. The subsequent sections describe the function of each table.

Table name	Short description of content
MAIN	Data of all samples for individual interferometers
ANTENNA	Antenna information
DATA_DESCRIPTION	Pointers to POLARIZATION and SPECTRAL_WINDOW entries
FEED	Feed (Frontend) related information
FIELD	Information on observed positions
FLAG_CMD	Flag information
OBSERVATION	General observation information
POINTING	Antenna pointing information
POLARIZATION	Polarization description information
PROCESSOR	Correlator information
SOURCE	Information on observed sources
SPECTRAL_WINDOW	Frequency/IF information
STATE	State information (mostly for single dish)
LOFAR_ANTENNA_FIELD	LOFAR Antenna Field information
LOFAR_ELEMENT_FAILURE	Time dependent failure information on tiles/dipoles
LOFAR_STATION	Station information
EPHEMERIDES	Ephemerides information
QUALITY_TIME_STATISTIC	Time related quality statistics
QUALITY_FREQUENCY_STATISTIC	Frequency related quality statistics
QUALITY_BASELINE_STATISTIC	Baseline related quality statistics
QUALITY_KIND_NAME	Index reference to what kinds of statistics are stored
HISTORY	History log of MS

2.2 MAIN table

The MAIN table of the MS contains the bulk of all data. For each interferometer and for each sample time The ordering is usually time-baseline, i.e. the MAIN table is divided in subsequent time-blocks. Within each time block there is an ordering based on the interferometer antenna pair. Each interferometer pair will appear only once per sample time. Auto- and crosscorrelations are usually mixed. The MAIN table links directly to many other tables through index numbers in several of its columns.

2.3 ANTENNA Table

The Antenna table contains information on each Antenna Field used in the observation.

2.4 DATA DESCRIPTION Table

This is a simple table that currently only contains identifiers for the SPECTRAL_WINDOW and POLARIZATION table.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

6 of 32

2.5 FEED table

This table gives information on the front-ends (Feeds) of the antennas used. It allows for time dependent beam and polarization properties through a TIME and INTERVAL column.

2.6 FIELD table

This table contains positional information on the object(s) that is pointed to by the telescope.

2.7 FLAG_CMD table

The FLAG_CMD table allows for additional flagging information in the MS, to be applied to the data in the MAIN table. For LOFAR this table is not being used, see the LOFAR_FULL_RES_FLAG column in the MAIN table instead.

2.8 OBSERVATION table

This table contains information on the project(s) and the scheduling. It is pointed to from the MAIN table by the OBSERVATION_ID column. It contains information related to the SAS/MAC/MoM Observation and Project and the other fields common to the different dataformats for LOFAR.

2.9 POINTING table

The POINTING table provides information on the actual pointing of an antenna, which may be time dependent. The structure and the fields are particularly useful for dish-based telescopes and interferometers.

2.10 POLARIZATION table

The POLARIZATION table provides information on the polarization properties of the receivers used in the observation. It refers to the polarization dependent fields in the FEED table through the CORR_PRODUCT field.

2.11 PROCESSOR table

This table gives some administrative information on the back-end processing. It can be used to store some information on the backend setup (e.g., the mode that was used).

2.12 SOURCE table

This table gives some information of the physical properties of the source(s) observed, as well as their use in the observing strategy (calibrator or not). For spectral line observations it contains information on the line rest frequency and the velocity of the object to be observed. This table is the only non-compulsory table we added to the LOFAR Measurement Set as it helps to clarify the goal of the observation.

2.13 SPECTRAL_WINDOW table

This table presents the frequency settings of the observation.

2.14 STATE table

This is mostly in use for single dish and may not be useful for LOFAR.

2.15 LOFAR_ANTENNA_FIELD table

This is a custom table for use by LOFAR only. It contains additional information specific to each antenna field. This table is required to support the mode where OBSERVATION.LOFAR_ANTENNA_SET is equal to HBA_JOINED because multiple antenna fields at the station are combined into a single Antenna Field.

2.16 LOFAR_ELEMENT_FAILURE table

This is a custom table for use by LOFAR only. It contains information about failure in time of the individual dipoles/tiles of each antenna field.

2.17 LOFAR_STATION table

This is a custom table for use by LOFAR only. It contains a list of all the stations that were present in the observation, mainly so the ANTENNA table can reference them with the purpose of identifying which antenna fields are part of which station.

2.18 EPHEMERIDES table

This is a table that is referenced in the MS 2.0 definition but never defined. It is currently not needed for LOFAR but added to this document for future extension and definition. This table will be needed if it becomes necessary to describe fields or objects being observed which move fast enough to not have a constant position on the sky during an observation. Note that a lot of objects like the Sun, Moon and planets already have ephemerides information available through the standard AIPS++/CASA DATA/EPHIMERIDES. This table is foreseen to provide ephemerides for objects not in that table, like satellites.

2.19 QUALITY_TIME_STATISTIC table

This is an optional table for storing time related quality statistics. It's currently only used in LOFAR MeasurementSets, but in principle it's not LOFAR specific.

2.20 QUALITY_FREQUENCY_STATISTIC table

This is an optional table for storing frequency related quality statistics. It's currently only used in LOFAR MeasurementSets, but in principle it's not LOFAR specific.

2.21 QUALITY_BASELINE_STATISTIC table

This is an optional table for storing baseline related quality statistics. It's currently only used in LOFAR MeasurementSets, but in principle it's not LOFAR specific.

2.22 QUALITY_KIND_NAME table

This is an optional table functioning as a lookup table for the kinds of data stored in the three above quality statistics tables. It's currently only used in LOFAR MeasurementSets, but in principle it's not LOFAR specific.

2.23 HISTORY table

This table allows HISTORY information to enter the MS so that one can trace what has happened to the MS since it was created. The on-line systems can add a creation statement here. It should mainly be filled and used by processes that alter the data in the MS after initial creation of the MS. Each step in the processing pipelines will store its parset file settings and other relevant data in this table.

**LOFAR/
USG/
Data Formats**

Doc.nr.:

Rev.:

Date:

Class.: Public

8 of 32

ASTRON

3 Detailed description of all MS tables

3.1 MAIN table

Name	Format	Units	Measure	Default	Comments
Keywords					
MS_VERSION	Float			2.0	MS version number
LOFAR_VERSION	Float			07.01	Version of this document
Key					
TIME	Double	s	EPOCH		Integration midpoint
ANTENNA1	Int				Pointers to ANTENNA table
ANTENNA2	Int				
FEED1	Int				Pointers to FEED_ID column in FEED table
FEED2	Int				
DATA_DESC_ID	Int				Pointer to DATA_DESCRIPTOR table
PROCESSOR_ID	Int				Pointer to PROCESSOR table
FIELD_ID	Int				Pointer to FIELD table
Non-key attributes					
UVW	Double (3)	m	UVW J2000		UVW coordinates at TIME_CENTROID in J2000
INTERVAL	Double	s			Sampling interval (this sample)
EXPOSURE	Double	s			Effective integration time
TIME_CENTROID	Double	s	EPOCH		Average time of sample
SCAN_NUMBER	Int			0	Scan number
ARRAY_ID	Int			0	Sub-array number
OBSERVATION_ID	Int			0	Pointer to OBSERVATION table
STATE_ID	Int			0	Pointer to State table
Data					
DATA	Complex (Nc, Nf)	Jy			Raw complex visibilities
CORRECTED_DATA	Complex (Nc, Nf)			0	Corrected complex visibilities
MODEL_DATA	Complex (Nc, Nf)			0	Complex visibilities of data model
(VIDEO_POINT)	Complex (Nc)				Video point
SIGMA	Float (Nc)			[1,1,1,1]	Estimated rms noise for single channel
WEIGHT	Float (Nc)				Weight for whole DATA matrix
WEIGHT_SPECTRUM	Float (Nc, Nf)				Weight for each channel separately
Flag information					
FLAG	Bool(Nc, Nf)				Cumulative data flags
FLAG_CATEGORY	Bool (Nc, Nf, Ncat)			Empty	Flag categories
FLAG_ROW	Bool			F	Row flag
LOFAR_FULL_RES_FLAG	uChar(Nf/8, Ntimeavg)				Full resolution flag before compression

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

9 of 32

Note that *Nc*= number of independent correlation signals (i.e, polarizations), *Nf*= number of frequency channels, and *Ncat*= number of flag categories (not used for LOFAR, so *Ncat*=1), Ntimeavg = is the number of timeslots averaged to a single timeslot.

MS VERSION The MeasurementSet format revision number, expressed as *major revision.minor revision*. This version is 2.0.

LOFAR VERSION The LOFAR MeasurementSet format revision number as defined by this document.

TIME This is the mid-point (not centroid) of data interval in UTC.

ANTENNA*n* This is the antenna number (≥ 0), and a direct index into the ANTENNA sub-table *rownr*.

FEED*n* This is the feed number (≥ 0). This points to the FEED_ID column in the FEED table.

DATA_DESC_ID Data description identifier (≥ 0), and a direct index into the DATA DESCRIPTION sub-table *rownr*.

PROCESSOR_ID Processor identifier (≥ 0), and a direct index into the PROCESSOR sub-table *rownr*.

FIELD_ID Field identifier (≥ 0), and a direct index into the FIELD sub-table *rownr*.

INTERVAL Data sampling interval. This is the nominal data interval and does not include the effects of bad data or partial integration.

EXPOSURE This is the effective data interval, including bad data and partial averaging. For now equal to INTERVAL.

SCAN_NUMBER Arbitrary scan number to identify data taken in the same logical scan. Not required to be unique. For LOFAR this is used to identify records in mosaic observations that should be grouped by pointing or time.

ARRAY_ID This is the subarray identifier (≥ 0), which identifies data in separate sub arrays. In previous versions of the LOFAR MS this was always 0, but from version 2.08 the actual Sub-Array Pointing number use in the Observation will be written in here.

OBSERVATION_ID This is the observation identifier (≥ 0), a direct index into the OBSERVATION sub-table *rownr*. This is not the same as the ObservationId that is used by MoM, SAS and LTA, which can be found in the OBSERVATION table.

STATE_ID This is the state identifier (≥ 0), and a direct index into the STATE sub-table *rownr*. Always 0 for LOFAR.

UVW *uvw* coordinates for the baseline from ANTENNA2 to ANTENNA1, i.e. the baseline is equal to the difference POSITION2 - POSITION1. The UVW given are for the TIME_CENTROID, and correspond in general to the reference type for the PHASE_DIR of the relevant field. I.e. J2000 if the phase reference direction is given in J2000 coordinates. However, any known reference is valid. Note that the choice of baseline direction and UVW definition (*W* towards source direction; *V* in plane through source and system's pole; *U* in direction of increasing longitude coordinate) also determines the sign of the phase of the recorded data.

DATA These are the measured visibilities in units of correlation coefficients. This contains the output of the correlator.

CORRECTED_DATA The visibilities corrected for instrumental effects (calibration), or even after subtraction of a local/global sky model. When not used, this contains an array of zero values.

MODEL_DATA The predicted visibilities of a local/global sky model. When not used, this contains an array of zero values.

VIDEO_POINT The video point for the spectrum, to allow for the full reverse FFT transform of the spectrum to the correlation function. Can be ignored for LOFAR.

SIGMA The estimated rms noise for a single channel, for each correlator. If we don't have this number, it will be set to 1.

SIGMA_SPECTRUM The estimated rms noise for each channel.

WEIGHT The weight for the whole data matrix for each correlator, as assigned by the correlator or processor. Will be set to the average weight in the WEIGHT_SPECTRUM column (without taking FLAGs into account).

WEIGHT_SPECTRUM The weight for each channel in the data matrix, as assigned by the correlator or processor. The weight spectrum should be used in preference to the WEIGHT, when available. For LOFAR the weight spectrum is determined in the correlator by the ratio of received packets per integration time and then updated adjusted for flagging when compressing in time or frequency.

FLAG An array of Boolean values with the same shape as DATA (see the DATA item above) representing the cumulative flags applying to this data matrix, as specified in FLAG CATEGORY. Data are flagged as bad if the FLAG array element is True.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

10 of 32

FLAG_CATEGORY An array of flag matrices with the same shape as DATA, but indexed by category. The category identifiers are specified by a keyword CATEGORY, containing an array of string identifiers, attached to the FLAG CATEGORY column and thus shared by all rows in the MeasurementSet. The cumulative effect of these flags is reflected in column FLAG. Data are flagged bad if the FLAG array element is True. This column is not used for LOFAR see LOFAR_FULL_RES_FLAG instead.

FLAG_ROW True if the entire row is flagged. Not used, so always set to False.

FULL_RES_FLAG These are an array of flags at the full resolution of the original uncompressed data containing what flags were generated before compression. This is needed to correct for bandwidth and time smearing due to compression. It contains bits. uChar is used for it, because Int is sensitive to endianness. It contains a 2-dim array with shape [(Nf+7)/8, Ntimeavg] where Ntimeavg is the number of time slots averaged to a single time slot. As for LOFAR all polarizations are flagged if one is flagged, no axis with Nc length is needed.

Note that in the last time slot of an averaged MS, FULL_RES_FLAG will contain True flags for in case of missing time slots in the original MS.

The column has two keywords defined with it:
NCHAN_AVG defines the number of channels averaged to one.
NTIME_AVG defines the number of time slots averaged to one. It is equal to the second axis of the shape.

Finally note that a FLAG value True does not mean that all corresponding FULL_RES_FLAG values are True. But when averaging further, FULL_RES_FLAG values will be set to True for FLAG values equal to True.

3.2 ANTENNA table

ANTENNA Table: Antenna Field Characteristics					
Name	Format	Units	Measure	Default	Comments
Data					
NAME	String			LOFAR	Unique antenna field names
LOFAR_STATION_ID	Int				Pointer to STATION table
TYPE	String			GROUND-BASED	Antenna Type
MOUNT	String			X-Y	Antenna Mounting
POSITION	Double (3)	m	POSITION		Antenna center of light position in ITRF or WGS84
OFFSET	Double (3)	m	POSITION	0	Axes offset of mount to FEED_REFERENCE point
DISH_DIAMETER	Double	m		0	No meaning for LOFAR
LOFAR_PHASE_REFERENCE	Double (3)	m	POSITION		Beam forming phase reference
Flag information					
FLAG_ROW	Bool			False	Row flag

Notes: This sub-table contains the global antenna field properties for each antenna field in the MS. It is indexed directly from MAIN via ANTENNAn.

NAME Antenna field name. This is either HBA, LBA, HBA0 or HBA1.
LOFAR_STATION_ID This is the station identifier (≥ 0), a direct index into the LOFAR_STATION sub-table *rownr*.
TYPE Antenna type. Reserved keywords include: (”GROUND-BASED” - conventional antennas; ”SPACE-BASED” - orbiting antennas; ”TRACKING-STN” - tracking stations). We will use “GROUND-BASED”.

LOFAR/ USG/ Data Formats

Doc.nr.:
Rev.:
Date:
Class.: Public

11 of 32

MOUNT The mount type of the antenna. Reserved keywords include: ("EQUATORIAL" - equatorial mount; "ALTAZ"- azimuth-elevation mount; "X-Y" - x-y mount; "SPACE-HALCA" - specific orientation model.). For LOFAR we will use "X-Y".

POSITION In a right-handed frame, X towards the intersection of the equator and the Greenwich meridian, Z towards the pole. The exact frame should be specified in the MEASURE_REFERENCE keyword (ITRF or WGS84) attached to this column. On traditional telescopes the reference point is the point on the azimuth or horizontal ascension axis closest to the elevation or declination axis. For LOFAR this is the effective centre of the collecting area of the antenna field, usually a weighted average of the positions of the individual antennas. It is needed because one calculates the (u,v,w) coordinates as differences of the center of light positions between antenna fields, not the phase reference positions.

OFFSET Axes offset of mount to feed reference point.

DISH_DIAMETER This is the nominal diameter of dish, as opposed to the effective diameter. This is not a constant for LOFAR. For LOFAR this is set at 0 meter to make it obvious that it should not be used.

FLAG_ROW This is the boolean flag to indicate the validity of this entry. Set to True for an invalid row. This does not imply any flagging of the data in MAIN, but is necessary as the ANTENNA index in MAIN points directly into the ANTENNA sub-table. Thus FLAG ROW can be used to delete an antenna entry without re-ordering the ANTENNA indices throughout the MS.

LOFAR_PHASE_REFERENCE This is the phase reference position, the location at which the antenna field tries to adjust the "phases" for all of the dipoles of an antenna field to be equal for the Sub Array Pointing direction.

Time variant LOFAR antenna field and station properties (e.g., switching individual dipoles in a field off during an observation) cannot be handled directly in the antenna table. This is only relevant when such a change leads to different antenna characteristics (e.g., center of light position). For the current version of the LOFAR Measurement Set the choice has been made to handle this though the LOFAR_ELEMENT_FAILURE table. This means that the effect such changes have on the center of light position are not reflected in the POSITION in this table.

3.3 DATA_DESCRIPTION Table

Name	Format	Units	Measure	Default	Comments
DATA_DESCRIPTION Table: Frequency/polarization characteristics					
Data					
SPECTRAL_WINDOW_ID	Int				Index into the SPECTRAL_WINDOW table
POLARIZATION_ID	Int				Index into the POLARIZATION table
Flag information					
FLAG_ROW	Bool				Row flag

Notes: This table defines the shape of the associated DATA array in MAIN, and is indexed directly by DATA_DESC_ID.

SPECTRAL_WINDOW_ID Spectral window identifier. It is a direct index into the SPECTRAL_WINDOW sub-table.

POLARIZATION_ID Polarization identifier (≥ 0). It is a direct index into the POLARIZATION sub-table.

FLAG ROW True if the row does not contain valid data; This does not imply flagging in MAIN.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

12 of 32

3.4 FEED table

FEED Table: Feed characteristics					
Name	Format	Units	Measure	Default	Comments
Key					
ANTENNA_ID	Int				
FEED_ID	Int			0	
SPECTRAL_WINDOW_ID	Int			-1	
TIME	Double	s	EPOCH		Interval midpoint
Data description					
NUM_RECEPTORS	Int			2	Number of receptors on this feed
Data					
BEAM_ID	Int			-1	Beam model
BEAM_OFFSET	Double (2, NUM_RECEPTORS)	rad	DIRECTION	0	Beam position offset (on sky but in antenna reference frame).
POLARIZATION_TYPE	String (NUM_RECEPTORS)			X, Y	Type of polarization to which a given RECEPTOR responds.
POL_RESPONSE	Complex (NUM RECEPTORS, NUM RECEPTORS)				Feed polarization response
POSITION	Double(3)	m	POSITION	0,0,0	Position of feed relativeto feed reference position for this antenna
RECEPTOR_ANGLE	Double (NUM RECEPTORS)	rad			The reference angle for polarisation.

Notes: A feed is a collecting element on an antenna, such as a single horn, that shares joint physical properties and makes sense to calibrate as a single entity. It is an abstraction of a generic antenna feed and is considered to have one or more RECEPTORs that respond to different polarization states. A FEED may have a time variable beam and polarization response. Feeds are numbered from 0 on each separate antenna for each SPECTRAL_WINDOW_ID. Consequently, FEED_ID should be non-zero only in the case of feed arrays, i.e. multiple, simultaneous beams on the sky at the same frequency and polarization. For LOFAR this describes the behaviour of the signal after having the signals from the individual dipoles beam formed within an antenna field. It does not describe the individual dipoles. For LOFAR this in essence describes a virtual device.

ANTENNA_ID Antenna field number, as indexed from ANTENNAn in MAIN.
FEED_ID Feed identifier, as indexed from FEEDn in MAIN.
SPECTRAL_WINDOW_ID Spectral window identifier: A value of -1 indicates the row is valid for all spectral windows.
TIME This value is the mid-point of time interval for which the feed parameters in this row are valid. The same Measure reference used for the TIME column in MAIN must be used.
INTERVAL Time interval for which the feed parameters in this row are valid.
NUM_RECEPTORS Number of receptors on this feed. See POLARIZATION_TYPE for further information.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

13 of 32

BEAM_ID Beam identifier. Points to an optional BEAM sub-table defining the primary beam and polarization response for this FEED. A value of -1 indicates that no associated beam response is defined.
BEAM_OFFSET Beam position offset, as defined on the sky but in the antenna reference frame.
POLARIZATION_TYPE Polarization type to which each receptor responds (e.g. "R","L","X" or "Y"). This is the receptor polarization type as recorded in the final correlated data (e.g. "RR"); i.e. as measured after all polarization combiners.
POL_RESPONSE Polarization response at the centre of the beam for this feed. Expressed in a linearly polarized basis (*ex,ey*) using the IEEE convention. For LOFAR this is (1,0),(0,1)
POSITION This is the offset of feed relative to the feed reference position for this antenna (see the ANTENNA sub-table). For LOFAR this is always (0,0,0)
RECEPTOR_ANGLE Polarization reference angle. Converts into parallactic angle in the sky domain.

3.5 FIELD table

<table>
<tr><th colspan="6">FIELD Table: Field positions for each source</th></tr>
<tr><th>Name</th><th>Format</th><th>Units</th><th>Measure</th><th>Default</th><th>Comments</th></tr>
<tr><td colspan="6">*Data*</td></tr>
<tr><td>NAME</td><td>String</td><td></td><td></td><td></td><td>Name of field</td></tr>
<tr><td>CODE</td><td>String</td><td></td><td></td><td></td><td>Special characteristics</td></tr>
<tr><td>TIME</td><td>Double</td><td>s</td><td>EPOCH</td><td></td><td>Time origin for the directions and rates</td></tr>
<tr><td>NUM_POLY</td><td>Int</td><td></td><td></td><td>0</td><td>Polynomial series order</td></tr>
<tr><td>DELAY_DIR</td><td>Double (2, NUM_POLY+1)</td><td>rad</td><td>DIRECTION</td><td></td><td>Direction of delay centre</td></tr>
<tr><td>PHASE_DIR</td><td>Double (2, NUM_POLY+1)</td><td>rad</td><td>DIRECTION</td><td></td><td>Direction of Phase centre</td></tr>
<tr><td>REFERENCE_DIR</td><td>Double (2, NUM_POLY+1)</td><td>rad</td><td>DIRECTION</td><td></td><td>Direction of Reference centre</td></tr>
<tr><td>SOURCE_ID</td><td>Int</td><td></td><td></td><td>-1</td><td>Index in SOURCE table</td></tr>
<tr><td>LOFAR_TILE_BEAM_DIR</td><td>Double (2, NUM_POLY+1)</td><td>rad</td><td>DIRECTION</td><td></td><td>Direction of Tile Beam</td></tr>
<tr><td colspan="6">*Flag information*</td></tr>
<tr><td>FLAG_ROW</td><td>Bool</td><td></td><td></td><td>False</td><td>Row flag</td></tr>
</table>

Notes: The FIELD table defines a field position on the sky. For interferometers, this is the correlated field position.

NAME Field name: user specified in proposal or MoM.
CODE Field code indicating special characteristics of the fields, user specified.
TIME Time reference for the directions and rates. Required to use the same TIME Measure reference as in MAIN. We use the start time of the observation.
NUM_POLY Series order for the * DIR columns. This can be used to describe time-variant behaviour of these direction columns. Time origin is given by the TIME field. Default value is 0 (no polynomial expansion).
DELAY_DIR Direction of delay centre: This can be expressed as a polynomial in time, but we will only give constant values. Final result converted to the defined Direction Measure type.
PHASE_DIR Direction of phase (fringe stopping) centre: This can be expressed as a polynomial in time, but we will only give constant values. Final result converted to the defined Direction Measure type.
REFERENCE_DIR Reference centre: This can be expressed as a polynomial in time, but we will only give constant values. Final result converted to the defined Direction Measure type. For interferometric data, this is the original correlated field centre, and may equal DELAY_DIR or PHASE_DIR.
SOURCE_ID Points to an entry in the (optional) SOURCE subtable, a value of −1 indicates there is no corresponding source defined.

LOFAR_TILE_BEAM_DIR Direction of Tile Beam centre: During multi beam observations, the tile beam can only point into one direction, which means it needs to be specified separately.
FLAG_ROW True if data in this row are invalid, else False. Does not imply flagging in MAIN.

When doing position mosaic, the number of entries in this table will equal the number of mosaic positions.

3.6 OBSERVATION table

OBSERVATION Table: Observation information					
Name	Format	Units	Measure	Default	Comments
Data					
TELESCOPE_NAME	String			LOFAR	
TIME_RANGE	Double(2)	s	EPOCH		Specified start/end times
OBSERVER	String			Observer	Name of observer
LOG	String (*)				Observing log
SCHEDULE_TYPE	String			LOFAR	
SCHEDULE	String (*)			[corrSchedule]	Project Schedule
PROJECT	String			NONE	Project identification
LOFAR_PROJECT_TITLE	String				Project description
LOFAR_PROJECT_PI	String				Primary Investigator
LOFAR_PROJECT_CO_I	String				Co Investigators
LOFAR_PROJECT_CONTACT	String				Contact Author
LOFAR_OBSERVATION_ID	String				Observation ID
LOFAR_OBSERVATION_START	Double	s	EPOCH		Observation start
LOFAR_OBSERVATION_END	Double	s	EPOCH		Observation end
LOFAR_OBSERVATION_FREQUENCY_MAX	Double	MHz			Maximum frequency
LOFAR_OBSERVATION_FREQUENCY_MIN	Double	MHz			Minimum frequency
LOFAR_OBSERVATION_FREQUENCY_CENTER	Double	MHz			Centre Frequency
LOFAR_SUB_ARRAY_POINTING	Int				SubArrayPointing
LOFAR_ANTENNA_SET	String				SAS AntennaSet
LOFAR_FILTER_SELECTION	String				SAS FilterSelection
LOFAR_CLOCK_FREQUENCY	Double	MHz			SAS Clock setting
LOFAR_NOF_BITS_PER_SAMPLE	Int				NrBitsPerSample
LOFAR_TARGET	String (*)				Single or list
LOFAR_SYSTEM_VERSION	String				As per ICD
LOFAR_PIPELINE_NAME	String				Pipeline identification
LOFAR_PIPELINE_VERSION	String				Pipeline version
LOFAR_FILENAME	String				As per ICD
LOFAR_FILETYPE	String			uv	As per ICD
LOFAR_FILEDATE	Double	s	EPOCH		As per ICD
RELEASE_DATE	Double	s	EPOCH		Target release date
Flag information					
FLAG_ROW	Bool			False	Row flag

Notes: This table contains information specifying the observing instrument or epoch. It is indexed directly from MAIN via OBSERVATION_ID.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.:   Public

15 of 32

TELESCOPE_NAME Telescope name ("LOFAR").

TIME_RANGE The start and end times of the overall observing period spanned by the actual recorded data in MAIN. Required to use the same TIME Measure reference as in MAIN.

OBSERVER The name(s) of the observer(s).

LOG The observing log, as supplied by the telescope or instrument.

SCHEDULE_TYPE The schedule type, with current reserved types ("VLBA-CRD", "VEX", "WSRT", "ATNF"). "LOFAR" for LOFAR.

SCHEDULE Unmodified schedule file, of the type specified, and as used by the instrument.

PROJECT Project code (e.g. "LEA1234")

LOFAR_PROJECT_TITLE Description if the project (e.g. "Polarization properties of 3C1234")

LOFAR_PROJECT_PI Project primary investigator

LOFAR_PROJECT_CO_ I comma separated list of co-investigators

LOFAR_PROJECT_CONTACT e-mail address of the contact author

LOFAR_OBSERVATION_ID Observation ID as used by SAS, MAC and MOM based on the SAS VIC tree number.

LOFAR_OBSERVATION_START Observation start time/date. Doesn't need to be the same as for this specific SubArrayPointing or MeasurmentSet.

LOFAR_OBSERVATION_END Observation end time/date. Doesn't need to be the same as for this specific SubArrayPointing or MeasurmentSet.

LOFAR_OBSERVATION_FREQUENCY_MAX Observation maximum frequency. Not for a single spectral window, but for the entire observation.

LOFAR_OBSERVATION_FREQUENCY_MIN Observation minimum frequency. Not for a single spectral window, but for the entire observation.

LOFAR_OBSERVATION_FREQUENCY_CENTER Observation center frequency. Not for a single spectral window, but for the entire observation.

LOFAR_SUB_ARRAY_POINTING Number of the SubArrayPointing used for this measurement within the SAS observation

LOFAR_ANTENNA_SET SAS AntennaSet setting (e.g. "LBA_SPARSE_INNER", "HBA_ONE")

LOFAR_FILTER_SELECTION SAS FilterSelection setting (e.g. "10-90 MHz")

LOFAR_CLOCK_FREQUENCY SAS ClockFrequency setting (e.g. "160 Mhz")

LOFAR_NOF_BITS_PER_SAMPLE SAS NrBitsPerSample setting: 4, 8 or 16

LOFAR_TARGET Single or list of targets. This should basically be a condensed version of the SOURCES table. Can be empty.

LOFAR_SYSTEM_VERSION Data processing system name and version number

LOFAR_PIPELINE_NAME Pipeline processing name. If there are multiple runs only the first one will be identified here, further processing details can still be found in the HISTORY table.

LOFAR_PIPELINE_VERSION Pipeline version

LOFAR_FILENAME File name

LOFAR_FILETYPE File type "uv" for MeasurementSets

LOFAR_FILEDATE File creation date

RELEASE_DATE Project release date. This is the date on which the data may become public if the project finishes as scheduled.

FLAG_ROW Row flag. True if data in this row is invalid, but does not imply any flagging in MAIN.

Notes: The OBSERVATION_NOF_STATIONS and OBSERVATION_STATIONS_LIST as used in the ICDs are not used here, as the information is readily available from the LOFAR_STATION table.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

16 of 32

3.7 POINTING table

Name	Format	Units	Measure	Default	Comments
POINTING Table: Antenna Pointing information					
Key					
ANTENNA_ID	Int				
TIME	Double	S	EPOCH		Interval midpoint
INTERVAL	Double	S			Interval duration
Data					
NAME	String				Pointing position descriptor
NUM_POLY	Int			0	Polynomial series order
TIME_ORIGIN	Double	S	EPOCH		Origin for polynomial.
DIRECTION	Double (2, NUM POLY+1)	Rad	DIRECTION		Antenna pointing direction (J2000)
TARGET	Double (2, NUM POLY+1)	Rad	DIRECTION		Target direction (J2000)
TRACKING	Bool			True	True if on-position

Notes: This table contains information concerning the primary pointing direction of each antenna as a function of time. Note that the pointing offsets for individual feeds on a given antenna are specified in the FEED sub-table with respect to this pointing direction.

ANTENNA_ID Antenna identifier, as specified by ANTENNAn in MAIN.
TIME This is the mid-point of the time interval for which the information in this row is valid. Required to use the same TIME Measure reference as in MAIN.
INTERVAL Time interval during which the information in this row is valid.
NAME Pointing direction name; user specified. Leave empty.
NUM_POLY Series order for the polynomial expressions in DIRECTION and POINTING OFFSET. Default value is 0 (no polynomial expansion)
TIME_ORIGIN Time origin for the polynomial expansions. Equal this to the start time of the observation.
DIRECTION Antenna pointing direction, optionally expressed as polynomial coefficients. The final result is interpreted as a Direction Measure using the specified Measure reference. Use RA, Dec in J2000.
TARGET Target pointing direction, optionally expressed as polynomial coefficients. The final result is interpreted as a Direction Measure using the specified Measure reference. This is the true expected position of the source, including all coordinate corrections such as precession, nutation etc. For now this is equal to DIRECTION.
TRACKING True if tracking the nominal pointing position.

3.8 POLARIZATION table

POLARIZATION Table: Polarization setup information					
Name	Format	Units	Measure	Default	Comments
Data description					
NUM_CORR	Int			4	Nr. Of correlation cross-products
Data					
CORR_TYPE	Int (NUM_CORR)				Polarization of correlation
CORR_PRODUCT	Int (2, NUM_CORR				Receptor cross products
Flag information					
FLAG_ROW	Bool			False	Row flag

Notes: This table defines the polarization labelling of the DATA array in MAIN, and is directly indexed from the DATA DESCRIPTION table via POLARIZATION ID.

NUM_CORR The number of correlation polarization products. For example, for (RR) this value would be 1, for (RR, LL) it would be 2, and for (XX, XY, YX, YY) it would be 4, etc.
CORR_TYPE An integer for each correlation product indicating the Stokes type as defined in the CASA Stokes class enumeration. For (XX, XY, YX, YY) this is (9,10,11,12).
CORR_PRODUCT Pair of integers for each correlation product, specifying the receptors from which the signal originated. The receptor polarization is defined in the POLARIZATION_TYPE column in the FEED table. An example would be (0,0), (0,1), (1,0), (1,1) to specify all correlations between two receptors.
FLAG ROW Row flag. True if the data in this row are not valid, but does not imply the flagging of any DATA in MAIN.

3.9 PROCESSOR table

PROCESSOR Table: Processor (backend) information					
Name	Format	Units	Measure	Default	Comments
Data					
TYPE	String			CORRELATOR	Processor type
SUB_TYPE	String			CEP	sub-type
TYPE_ID	Int			0	Not used for LOFAR
MODE_ID	Int			0	Not used for LOFAR
Flag information					
FLAG_ROW	Bool			False	Row flag

Notes: This table holds summary information for the back-end processing device used to generate the basic data in the MAIN table. Such devices include correlators, radiometers, spectrometers, pulsar-timers, amongst others.

TYPE Processor type; reserved keywords include ("CORRELATOR" - interferometric correlator; "SPECTROMETER" - single-dish correlator; "RADIOMETER" - generic detector/integrator; "PULSAR-TIMER" – pulsar timing device). For LOFAR this will be 'CORRELATOR'
SUB_TYPE Processor sub-type, e.g. "GBT" or "JIVE". For LOFAR 'CEP' will be used.
TYPE_ID Index used in a specialized sub-table named as subtype type, which contains time-independent processor information applicable to the current data record (e.g. a JIVE CORRELATOR sub-table). Time-dependent information for each device family is contained in other tables, dependent on the device type.
MODE ID Index used in a specialized sub-table named as subtype type mode, containing information on the processor mode applicable to the current data record. (e.g. a GBT SPECTROMETER MODE table).
FLAG_ROW Row flag. True if data in the row are not valid, but does not imply flagging in MAIN.

LOFAR/ USG/ Data Formats

Doc.nr.:
Rev.:
Date:
Class.: Public

18 of 32

3.10 SOURCE table (Optional)

SOURCE Table: Source information					
Name	Format	Units	Measure	Default	Comments
Key					
SOURCE_ID	Int				
TIME	Double	s	EPOCH		Midpoint of interval
INTERVAL	Double	s			Duration of interval
SPECTRAL_WINDOW_ID	Int				
Data description					
NUM_LINES	Int				Number of Spectral lines
Data					
NAME	String				Name of source during observation
CALIBRATION_GROUP	Int				Group number for calibration purposes
CODE	String				Special characteristics of source, e.g. Bandpass calibrator
DIRECTION	Double(2)	rad	DIRECTION		Direction of Source
PROPER_MOTION	Double(2)	rad/s		0	

Notes: This table contains time-variable source information, optionally associated with a given FIELD ID.

SOURCE_ID Source identifier (≥ 0), as specified in the FIELD sub-table.
TIME This is the mid-point of the time interval for which the data in this row are valid. Required to use the same TIME Measure reference as in MAIN.
INTERVAL Time interval for which the data in this row are valid.
SPECTRAL_WINDOW_ID Spectral window identifier: A -1 indicates that the row is valid for all spectral windows.
NUM_LINES Number of spectral line transitions associated with this source and SPECTRAL_WINDOW_ID combination.
NAME Source name: User specified.
CALIBRATION_GROUP Calibration group number to which this source belongs; user specified.
CODE Source code, used to describe any special characteristics of the source, such as the nature of a calibrator. Reserved keyword, including ("BANDPASS CAL").
DIRECTION Source direction at this TIME.
PROPER MOTION Source proper motion at this TIME.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

19 of 32

3.11 SPECTRAL_WINDOW table

SPECTRAL WINDOW Table: Frequency setup information					
Name	Format	Units	Measure	Default	Comments
Data description					
NUM_CHAN	Int			256	Number of Spectral channels
Data					
NAME	String			SB-x	Subband name/identifier
REF_FREQUENCY	Double	Hz	FREQUENCY		Reference Frequency
CHAN_FREQ	Double (NUM_CHAN)	Hz	FREQUENCY		Centre frequencies for each channel in the data matrix.
CHAN_WIDTH	Double (NUM_CHAN)	Hz			Channel width for each channel in the data matrix.
MEAS_FREQ_REF	Int			5	FREQUENCY Measure ref.
EFFECTIVE_BW	Double (NUM_CHAN)	Hz			The effective noise bandwidth of each spectral channel
RESOLUTION	Double (NUM_CHAN)	Hz			The effective spectral resolution of each channel
TOTAL_BANDWIDTH	Double	Hz			Total bandwidth for this Spectral window
NET_SIDEBAND	Int			0	Upper or lower
IF_CONV_CHAIN	Int			0	The IF conversion chain
FREQ_GROUP	Int			-1	Not defined for LOFAR
FREQ_GROUP_NAME	String				Same as FILTER_SELECTION in OBSERVATION table
Flag information					
FLAG_ROW	Bool			False	Row flag

Notes: This table describes properties for each defined spectral window. A spectral window is both a frequency label for the associated DATA array in MAIN, but also represents a generic frequency conversion chain that shares joint physical properties and makes sense to calibrate as a single entity.

NUM_CHAN Number of spectral channels (must be equal throughout the MS). For LOFAR this number always is *256*.
NAME Spectral window name: For LOFAR we use SBx where x is a number starting from 0 that indicates the subband number used. A LOFAR MS may have multiple subbands, but a single subband can only be in a single MS.
REF_FREQUENCY The reference frequency. This is the frequency representative of this spectral window, usually the sky frequency corresponding to the DC edge of the baseband. Used by the calibration system if a fixed scaling frequency is required or in algorithms to identify the observing band. Note: For the WSRT this field gives the *midband* sky frequency of the IF band (NOT the frequency corresponding to the DC edge of the IF band, as stated in the MS2 definition). It is also the frequency of channel number (NUM_CHAN/2) + 1 in the array given in field CHAN_FREQ.
CHAN_FREQ Centre frequencies for each channel in the data matrix. Note that the channel frequencies may be in ascending or descending frequency order.

**LOFAR/
USG/
Data Formats**

Doc.nr.:

Rev.:

Date:

Class.: Public

20 of 32

CHAN_WIDTH Nominal channel width of each spectral channel. Although these can be derived from CHAN_FREQ by differencing, it is more efficient to keep a separate reference to this information.

MEAS_FREQ_REF Frequency Measure reference for CHAN_FREQ. This allows a row-based reference for this column in order to optimize the choice of Measure reference when Doppler tracking is used. Modified only by the MS access code. The number is the enum value as given in the AIPS++ class MFrequency (5 is TOPO).

EFFECTIVE_BW The effective bandwidth of each spectral channel. Usually equals CHAN_WIDTH.

RESOLUTION The effective spectral resolution of each channel (may be different from CHAN_WIDTH if some form of tapering has been applied).

TOTAL_BANDWIDTH The total bandwidth for this spectral window.

NET_SIDEBAND The net sideband (upper or lower) for this spectral window.

IF_CONV_CHAIN Site specific identification of the electronic signal path for the case of multiple (simultaneous) Ifs (e.g. VLA: AC=0, BD=1, ATCA: Freq1=0, Freq2=1). Not relevant for LOFAR and always 0.

FREQ_GROUP A frequency group identifier to which the spectral window belongs. This is used to associate spectral windows for joint calibration purposes. For LOFAR this is not defined.

FREQ_GROUP_NAME The frequency group name. For LOFAR this contains the same value as the LOFAR_FILTER_SELECTION in the OBSERVATION table.

FLAG_ROW True if the row does not contain valid data.

3.12 STATE table

STATE Table: State information					
Name	Format	Units	Measure	Default	Comments
Data					
SIG	Bool			True	Signal
REF	Bool			False	Reference
CAL	Double	K		0	Noise calibration
LOAD	Double	K		0	Load temperature
SUB_SCAN	Int			0	Sub-scan number
OBS_MODE	String			Empty	Observing mode
Flag information					
FLAG_ROW	Bool			False	Row flag

Notes: This table defines the state parameters for a particular data record as they refer to external loads, calibration sources or references, and also characterizes the observing mode of the data record, as an aid to defining the scheduling heuristics. It is indexed directly via STATE_ID in MAIN.

SIG True if the source signal is being observed.

REF True for a reference phase.

CAL Noise calibration temperature (zero if not added).

LOAD Load temperature (zero if no load).

SUB_SCAN Sub-scan number (≥ 0), relative to the SCAN_NUMBER in MAIN. Used to identify observing sequences.

OBS_MODE Observing mode: defined by a set of reserved keywords characterizing the current observing mode (e.g."OFF-SPECTRUM"). It is used to define the schedule strategy.

FLAG_ROW True if the row does not contain valid data. Does not imply flagging in MAIN.

This table is compulsory, but has no direct relevance to LOFAR. Its entries will not be used in any post-processing tool as far as I am aware. We may use it eventually for simple calibration information (e.g. LOAD temperature), but currently it's not has no use within LOFAR.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

21 of 32

3.13 LOFAR_ANTENNA_FIELD table

LOFAR_DIPOLE Table: Dipole information					
Name	Format	Units	Measure	Default	Comments
Data					
NAME	String				Either HBA, LBA, HBA0 or HBA1
ANTENNA_ID	Int				Pointer to ANTENNA table
POSITION	Double(3)	m	POSITION		Position of antenna field
COORDINATE_SYSTEM	Double(3,3)	m	DIRECTION		P,Q,R
TILE_ROTATION	Double	rad			
TILE_ELEMENT_OFFSET	Double(3,16)	m	POSITION		
ELEMENT_OFFSET	Double(3, nelements)		POSITION		Offset per element
ELEMENT_FLAG	Bool(2, nelements)				Flag per element
Flag information					
FLAG_ROW	Bool			False	Row flag

Notes: This is a custom table for use by LOFAR only. It contains information about the one or more antenna fields that form each Antenna Field. It uses the ANTENNA_ID to specify to which Antenna Field the antenna field belongs.

NAME Name of the antenna field, currently either HBA, LBA, HBA0 or HBA1.
ANTENNA_ID Index into the ANTENNA table to show to which antenna field this dipole belongs.
POSITION Position of the antenna field in absolute ITRF coordinates.
COORDINATE_SYSTEM (cartesian) direction vectors in ITRF (or measure defined) describing the local field coordinate system (see "LOFAR reference plane and reference direction", Michiel Brentjens).
This is the up direction of the antennas, and is supposed to be the normal direction to the antenna field plane. Note that in general this is not the zenith direction, and can be many degrees away from zenith for some antenna fields. You need to know this direction in order to do the antenna beam pattern directions. This also contains the information needed to determine the +X and +Y polarization directions.
TILE_ROTATION This is the angl at which the antenna pattern is pointing, in the ground plane of the antenna field. This affects where the HBA tile grating lobes will appear on the sky among other things. In principle this is redundant if all tile/dipole positions are known, but that's non-trivial and therefore this is included as a useful reference. NB. Only filled in for antenna fields of type HBA.
TILE_ELEMENT_OFFSET Position offsets in ITRF (or measure defined) of the dual dipole elements inside a tile, with respect to the centre of the tile. NB. Only valid for antenna fields of type HBA.
ELEMENT_OFFSET Relative offsets of each element from POSITION.
ELEMENT_FLAG Two flags for each row in ELEMENT_OFFSET identifying if a dipole (X or Y) in an element is operational.
FLAG_ROW True if the row does not contain valid data. Does not imply flagging in MAIN.

LOFAR/ USG/ Data Formats

Doc.nr.:
Rev.:
Date:
Class.: Public

22 of 32

3.14 LOFAR_ELEMENT_FAILURE table

<table>
<tr><th colspan="6">LOFAR_ELEMENT_FAILURE Table: Dipole/tile failure information</th></tr>
<tr><th>Name</th><th>Format</th><th>Units</th><th>Measure</th><th>Default</th><th>Comments</th></tr>
<tr><td colspan="6">Data</td></tr>
<tr><td>ANTENNA_FIELD_ID</td><td>Int</td><td></td><td></td><td></td><td>Pointer to LOFAR_ANTENNA_FIELD table</td></tr>
<tr><td>ELEMENT_INDEX</td><td>Int</td><td></td><td></td><td></td><td>Pointer to ELEMENT_OFFSET array</td></tr>
<tr><td>TIME</td><td>Double</td><td>s</td><td>EPOCH</td><td></td><td>Time of failure</td></tr>
</table>

Notes: This is a custom table for use by LOFAR only. It contains information about the failure of separate dipoles/tiles of each antenna field. It uses the ELEMENT_ID to specify to which dipole/tile failed.

ANTENNA_FIELD_ID Index into the LOFAR_ANTENNA_FIELD table to show to which antenna field this element belongs.
ELEMENT_INDEX Index of the row in the ELEMENT_OFFSET array in the LOFAR_ANTENNA_FIELD table to identify which dipole/tile failed.
TIME Time when the failure was detected. Currently updated in 15 second intervals.

3.15 LOFAR_STATION table

<table>
<tr><th colspan="6">LOFAR_STATION Table: Station information</th></tr>
<tr><th>Name</th><th>Format</th><th>Units</th><th>Measure</th><th>Default</th><th>Comments</th></tr>
<tr><td colspan="6">Data</td></tr>
<tr><td>NAME</td><td>String</td><td></td><td></td><td></td><td>Name of the station</td></tr>
<tr><td>CLOCK_ID</td><td>Int</td><td></td><td></td><td></td><td>Index of shared clock</td></tr>
<tr><td colspan="6">Flag information</td></tr>
<tr><td>FLAG_ROW</td><td>Bool</td><td></td><td></td><td>False</td><td>Row flag</td></tr>
</table>

Notes: This is a custom table for use by LOFAR only. It contains a list of all the stations that were present in the observation, mainly so the ANTENNA table can reference with the purpose of identifying which antenna fields are part of which station. It is indexed directly from ANTENNA through STATION_ID. It currently only contains a few fields, but will probably be extended in the future.

NAME Name of the station. (e.g. "CS103", "DE602")
CLOCK_ID Number identifying a clock signal. It should be the same number for stations sharing a clock signal. Currently stations in the LOFAR Superterp share a clock.
FLAG_ROW True if the row does not contain valid data. Does not imply flagging in MAIN.

**LOFAR/
USG/
Data Formats**

Doc.nr.:

Rev.:

Date:

Class.: Public

23 of 32

3.16 QUALITY tables

The following four tables form a unit describing quality statistics information. This describes the general motivation and requirements behind these tables. This is largely a direct copy from the original proposal by A.Offringa.

3.16.1 Motivation

Calculating statistics over an entire measurement set requires traversing the whole main table. This is expensive, and some statistics might be unavailable after averaging or other processing. Therefore, it would be nice to store an overview of the statistics in a sub-table of the measurement set. The idea is that these tables are created and filled during flagging.
A particular use-case for these sub-tables is the LOFAR EoR diagnosis database (LEDDB) under construction by Oscar Martinez Rubi. The LEDDB will contain the statistics of all observations in the cluster. There are scripts periodically checking changes in a measurement set on the hard disks of the cluster, and will update the database if such changes are found. This already works well for the "instrument" sub-table. I can imagine the LOFAR observatory wants similar behaviour at some point.

3.16.2 Requirements

The main purpose of the table is getting a quick overview that allows some basic analysis of the RFI situation and the data quality. It should remain small, because all information is redundant, and we don't want to enlarge a set too much. For higher resolution statistics, one should refer to the original raw data. We want at least to have the following statistics:
- RFI ratios and data variances as a function of frequency at highest resolution (per channel)
- Per sub-band, RFI ratios and data variances as a function of time, at about minute scale resolution or higher if possible (currently stored at highest resolution)
- Per sub-band, RFI ratios and data variances per baseline.

These statistics would respectively have in a common single sub-band of 8 hours a size of:
- 256 numbers for a single frequency statistic
- – (8 x 60) = 480 number for a single time statistic
- with 50 stations, 0.5 x 50 x 51 = 1275 numbers for a single baseline statistic

This sums to around 2000 numbers (per polarization per statistic). Including some overhead, this would become around one megabyte per subband, and is quite acceptable. We need a few more statistics to allow combining ranges.

3.16.3 Implementation

Each .._STATISTIC table has one or more columns that defines the centre of the range. The assumption is that centroid or start and end ranges are not required for superficial statistics. The columns are:
- TIME for the QUALITY_TIME_STATISTIC table, with the same type and format as the TIME column in the main table.
- ANTENNA1 and ANTENNA2 for the QUALITY_BASELINE_STATISTIC table, with the same type as their same-named companian columns in het main table. These index the ANTENNA subtable.
- FREQUENCY for all tables, with the same type as the frequency column in the SPECTRAL_WINDOW sub-table and given in Hertz. For QUALITY_FREQUENCE_STATISTIC, this is the only range column.

The rational for adding a frequency range to all tables, is to allow combining several sub-band tables to one, without losing statistics resolution. Moreover, especially for the time statistics table, it can be useful to have a slightly higher frequency resolution (e.g. 4 steps / sub-band) for flagging in a second iteration, as I have done a few times. In the common situation of statistics for one sub-band, this column will normally contain the central frequency of the sub-band.

One thing that these statistics can not provide at this point, is per baseline (or antenna) statistics over time. Although these can be very useful for analysis, both dimensions are generally large. Therefore, if such statistics are required, it would be best to use the main table.

By default it would be useful to have the RFI ratio, the variance and the differential variance available. Note that calculating the total variance of two parts with only the variance of the two parts is not possible, which implies that one cannot calculate the total variance of an observation given the variances of the sub-bands, without traversing the main table. Therefore, I suggest not to store the Variance statistics by default, but the Count, Sum, and SumP2 statistics (and DCount, DSum and DSumP2 for the differential statistics), which do allow sub-bands to be combined.

From these, the variance can be calculated with: $(SumP2 - Sum*Sum/Count) / (Count-1.0)$
Similar for the differential variance: $(DSumP2 - DSum*DSum/DCount) / (DCount-1.0)$

Note that this is for one complex component – thus first evaluate the formula with real values, then with imaginary. Having a mean and variance also implies that the SNR can be calculated.

The statistics should always describe the entire measurement set, not a specific selection of the set. This also implies that if the set contains multiple pointings, beams or bands, the statistics are over all of them. Only in that case, the statistics might be non-uniformly gridded. However, the statistics should not include auto-correlations, except for the QUALITY_BASELINE_STATISTIC table, which should include auto-correlations. Always all the polarizations from the main table are described by the statistics.

A given quality kind for a given cut (time, frequence or baseline) is available if and only if
  A. the QUALITY_KIND_NAME table exists and contains the name of the statistic
  B. theQUALITY_X_STATISTIC table exist, with 'X' the desired cut through the data and
  C. the QUALITY_X_STATISTIC table contains at least one row for the corresponding kind.

New statistics, which are not yet mentioned in this document, can be added to the QUALITY_KIND_NAME table. Possibly, it might be interesting to specify a QUALITY_BASELINE_TIME_STATISTIC table, which should be the multi-dimensional combination of the QUALITY_BASELINE_STATISTIC and QUALITY_TIME_STATISTIC tables, but at this point there's no use for this yet. A smaller alternative is to have a QUALITY_ANTENNA_TIME_STATISTIC, which can summarize the (cross-correlation) statistics as a function of time per antenna. Such a table would be around 20 x larger compared to QUALITY_BASELINE_STATISTIC, around 8 MB for the common observation mentioned earlier.

3.17 QUALITY_TIME_STATISTIC table (Optional)

QUALITY_TIME_STATISTIC Table: Time related quality information					
Name	Format	Units	Measure	Default	Comments
Data					
TIME	Double	s	EPOCH		Time point
FREQUENCY	Double	Hz			Frequency point
KIND	Int				Pointer to QUALITY_KIND_STATISTIC table
VALUE	Complex (NUM_CORR)				Statistic values of KIND

LOFAR/
USG/
Data Formats

Doc.nr.:
Rev.:
Date:
Class.: Public

25 of 32

Notes: This is an optional table for storing time related quality statistics. It's currently only used in LOFAR MeasurementSets, but in principle it's not LOFAR specific.

TIME Time point for VALUE.
FREQUENCY Frequency point for VALUE.
KIND Reference to a QUALITY_KIND_NAME table entry.
VALUE Array of Complex values of NUM_CORR length. It contains statistical values of the type referenced by KIND.

3.18 QUALITY_FREQUENCY_STATISTIC table (Optional)

QUALITY_FREQUENCY_STATISTIC Table: FREQUENCY related quality information					
Name	Format	Units	Measure	Default	Comments
Data					
FREQUENCY	Double	Hz			Frequency point
KIND	Int				Pointer to QUALITY_KIND_STATISTIC table
VALUE	Complex (NUM_CORR)				Statistic values of KIND

Notes: This is an optional table for storing frequency related quality statistics. It's currently only used in LOFAR MeasurementSets, but in principle it's not LOFAR specific.

FREQUENCY Frequency point for VALUE.
KIND Reference to a QUALITY_KIND_NAME table entry.
VALUE Array of Complex values of NUM_CORR length. It contains statistical values of the type referenced by KIND.

3.19 QUALITY_BASELINE_STATISTIC table (Optional)

QUALITY_BASELINE_STATISTIC Table: Baseline related quality information					
Name	Format	Units	Measure	Default	Comments
Data					
ANTENNA1	Int				Pointers to ANTENNA table
ANTENNA2	Int				
FREQUENCY	Double	Hz			Frequency point
KIND	Int				Pointer to QUALITY_KIND_STATISTIC table
VALUE	Complex (NUM_CORR)				Statistic values of KIND

Notes: This is an optional table for storing baseline related quality statistics. It's currently only used in LOFAR MeasurementSets, but in principle it's not LOFAR specific.

ANTENNA*n* This is the antenna number (≥ 0), and a direct index into the ANTENNA sub-table *rownr*.
FREQUENCY Frequency point for VALUE.
KIND Reference to a QUALITY_KIND_NAME table entry.
VALUE Array of Complex values of NUM_CORR length. It contains statistical values of the type referenced by KIND.

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

26 of 32

3.20 QUALITY_KIND_NAME table (Optional)

QUALITY_KIND_NAME Table: Kind name lookup information					
Name	Format	Units	Measure	Default	Comments
Data					
KIND	Int				Index used in the other quality tables
NAME	String				Name if the quality parameter

Notes: This is an optional table functioning as a lookup table for the kinds of data stored in the three above quality statistics tables. It's currently only used in LOFAR MeasurementSets, but in principle it's not LOFAR specific.

KIND Index used to identify the parameter in the other QUALITY tables.
NAME Name of the quality parameter.

Currently the following quality parameters are defined:

The statistic represent respectively:
–   Count: The number of unflagged samples.
–   Sum: The sum of the unflagged data.
–   Mean: The mean of the unflagged data (Sum / Count).
–   RFICount: The number of flagged samples due to RFI. If possible, they should denote true RFI samples, not samples that are lost due to failing stations, border channel issues, etc.
–   RFISum: The sum of the flagged samples due to RFI.
–   RFIMean: The mean of the RFI (RFISum / RFICount).
–   RFIRatio:RFICount/(Count+RFICount).
–   FlaggedCount: The total number of flagged samples. This does include any kind of flagged samples. "NaN" / "inf" values should also be counted as being flagged. Count + FlaggedCount = Total nr samples in the set.
–   FlaggedRatio:FlaggedCount/(FlaggedCount+Count).
–   SumP2: The sum of squares. Together with the mean and the count, this provides enough information to calculate the variance. This easily allows combining ranges.
–   SumP3: The sum of the third power, see SumP4 for why.
–   SumP4: The sum of the fourth power. Together with the sum, sum of squares and sums of third power, this provides enough information to calculate the kurtosis and the variance of the variance. While the kurtosis and/or variance of variance can be stored directly, storing them as SumP4 allows to combine ranges easily.
–   Variance: The variance: $\sigma^2 = <(x - mean)^2>$.
–   VarianceOfVariance: The variance of the variance.
–   Skewness: The skewness (third standardized moment).
–   Kurtosis: The kurtosis, using the most common definition (i.e., normalized so that a kurtosis of Gausian distribution is zero).
–   SignalToNoise:Thesignaltonoise.
–   DSum,DMean,DSumP2,DSumP3,DSumP4,DVariance,DVarianceOfVariance,DCount: these statistics starting with 'D' are the same as their counterpart without the 'D', but are differential. They can be calculated in different ways, but they should quantize only the noise, thus contain as few contribution of celestial sources as possible. The most trivial option is to calculate the statistics on the difference between adjacent channels (on high res).
–   BadSolutionCount: as RFICount, but for bad solutions. This count should not include RFI marked samples. Count – BadSolutionCount = number of good solutions.
–   BadSolutionRatio:Count/BadSolutionCount.
–   CorrectCount, CorrectedMean, CorrectedSum, CorrectedSumP2: As their name counterparts, but for the CORRECTED_DATA column.
–   FTSum: The Fourier transform of the data without fringe stopping (thus phase centred on the NCP). Flagged samples should not be FTed. Only applicable in time and frequency direction.

LOFAR/ USG/ Data Formats

Doc.nr.:
Rev.:
Date:
Class.:   Public

27 of 32

Data should be uniformly gridded as the other statistics, and the time/frequency range value specifies the 1/time and 1/frequency values.
– FTMean:FTSum/Count

The RFICount, RFIRatio, FlaggedCount, FlaggedRatio, Count, DCount, BadSolutionCount, BadSolutionRatio and CorrectedCount are real valued, alle others are complex. The statistics of complex variables are independently calculated. For example, the complex sum of squares is: "sum {x in X} Real(x)^2 + i * imag(x)^2", and not "sum {x in X} x^2".

3.21 HISTORY table

HISTORY Table: History information					
Name	Format	Units	Measure	Default	Comments
Key					
Time	Double	s	EPOCH		Time-stamp for message
OBSERVATION_ID	Int				Points to OBSERVATION table
Data					
MESSAGE	String				Log message
PRIORITY	String				Message priority
ORIGIN	String				Code origin
OBJECT_ID	String				Originating ObjectID
APPLICATION	String				Application name
CLI_COMMAND	String(*)				CLI command sequence
APP_PARAMS	String(*)				Application parameters

Notes: This sub-table contants associated history information for the MS.

TIME Time-stamp for the history record. Required to have the same TIME Measure reference as used in MAIN.
OBSERVATION_ID Observation identifier. This is an index into the OBSERVATION table.
MESSAGE Log message.
PRIORITY Message priority with allowed types: DEBUGGING, WARN, NORMAL, SEVERE
ORIGIN Source code origin from which the message originated. Contains version/revision number.
OBJECT_ID Origination ObjectID if available, else blank.
APPLICATION Application name.
CLI_COMMAND CLI (Command Line Interface) command sequence invoking the application.
APP_PARAMS Application parameter values. For LOFAR this is an array of strings with the parset key-value pair values.

LOFAR/ USG/ Data Formats

Doc.nr.:
Rev.:
Date:
Class.: Public

28 of 32

Schema of the LOFAR MS

The figure below shows a graphical representation of the LOFAR Measurement Set. All subtables and columns in the MS are presented.
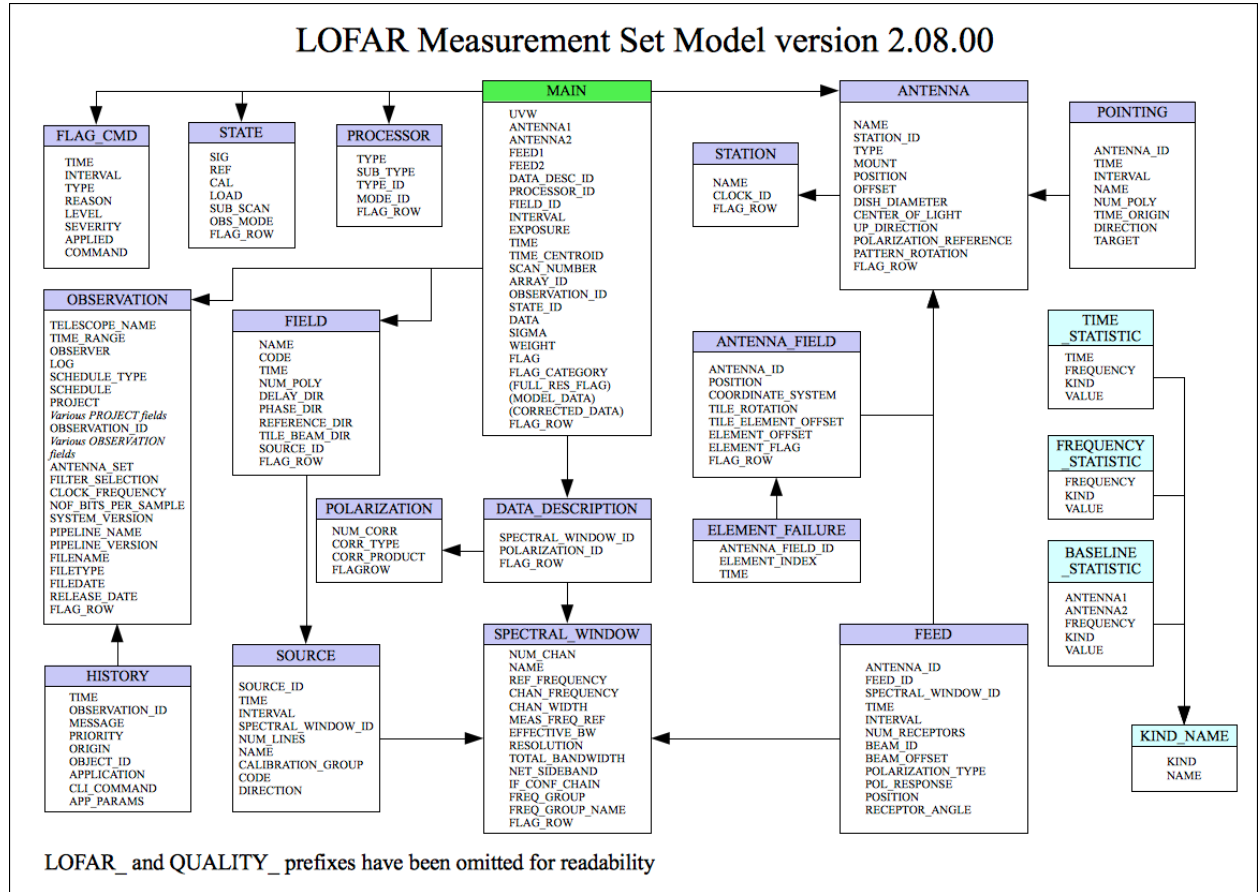

Figure 1: Schema of the LOFAR Measurement Set

LOFAR/
USG/
Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

29 of 32

Appendix A: Lofar Storage Manager (LofarStMan)

A.1 Introduction

For the 50-station LOFAR instrument the data rate for a single subband is about 10 MB/s. Because multiple subbands are kept on the same storage node, it is essential that the data can be dumped to disk as fast as possible, thus with hardly any overhead. On the other hand the data should be written in the standard casacore MeasurementSet format to be able to inspect and process them with the standard toolset.

Writing a MeasurementSet in the native way incurs overhead. One reason is that under the hood buffered IO is used; another reason is that several meta data columns are written that are constant for the LOFAR case.

MeasurementSets are handled by the casacore Table System that provides a way to use a dedicated storage manager. When accessing a table, the storage manager can be loaded dynamically provided that the shared library containing the code can be found in the library path.

This feature is used by the LOFAR data writer. It creates the MeasurementSet such that the Table System knows it has to use the special LofarStMan storage manager to understand the file containing the data that are written directly to disk.

A.2 LOFAR MeasurementSet creation

The data writer creates the MeasurementSet in the standard way, but binds the columns of the main table to the LofarStMan storage manager. It fills all subtables and creates a small file *table.f0meta* in the main table directory containing information like antenna numbers, time, endianness, etc.. It also contains a version number making it possible to evolve the format. The meta file is written in the casacore AipsIO format.

After this initialisation phase the MeasurementSet is closed and the main data file *table.f0data* is created. Data are written to that file when received from the online system. If possible O_DIRECT is used to avoid the kernel file cache overhead.

In version 1 and 2 each time slot in the file contains three blocks of data:
1. A 32-bit signed integer sequence number. It is used to derive the time.
2. Per channel the number of samples (16-bit unsigned integer) used by the correlator. It is used to derive the weights and flags.
3. The data array (ncorr,nchan) as single precision complex numbers.

The meta file tells how each block is aligned. Usually this is on 512 bytes because that is required to make use of the O_DIRECT option.

There can be gaps in the sequence numbers, thus time slots might be missing.
In case of a crash, the data are always fine. I.e., no indices or so have to be updated as would be the case if the native table format or a system like HDF5 is used.

Note that LofarStMan is only used for the raw data. The first NDPPP processing step will create a MeasurementSet using the standard storage managers.

	LOFAR/ USG/ Data Formats	Doc.nr.:	
		Rev.:	
		Date:	
		Class.:	Public

A.3 Accessing an existing LOFAR MeasurementSet

When an existing LOFAR MeasuementSet is opened, the Table System detects that the LofarStMan storage manager is needed and will load it dynamically from the shared library with that name. First a register function is called to make the storage manager known. Thereafter all accesses to the data are done through LofarStMan.
LofarStMan calculates the number or rows in the data file and reports it back to the Table System.

On 64-bit systems the data file is memory-mapped. In this way the system takes care of caching if random access is done. The memory space of 32-bit systems is too small for mmap, so buffered IO is used instead.
The columns handled by LofarStMan and their values are given in the following table. All columns except DATA are readonly. Note that bytes are swapped if the endianness requires it.

TIME	starttime + (seqnr + 0.5)*interval
ANTENNA1	from meta file
ANTENNA2	from meta file
FEED1	0
FEED2	0
DATA_DESC_ID	0
PROCESSOR_ID	0
FIELD_ID	0
INTERVAL	interval from meta file
EXPOSURE	interval from meta file
TIME_CENTROID	same as TIME
SCAN_NUMBER	0
ARRAY_ID	0
OBSERVATION_ID	0
STATE_ID	0
UVW	calculated on the fly using FIELD and ANTENNA subtables (for phase reference direction and antenna positions)
DATA	from data file
SIGMA	1
WEIGHT	1
WEIGHT_SPECTRUM	nr of samples / nominal nr of samples
FLAG	nr of samples == 0
FLAG_CATEGORY	empty
FLAG_ROW	false

A MeasurementSet stored with LofarStMan has some special characteristics:
- All columns are readonly with the exception of the DATA column.

	LOFAR/ USG/ Data Formats	Doc.nr.:
		Rev.:
		Date:
		Class.: Public

-   It is not possible to add or remove rows.
-   It is possible to remove a column.
-   A column can be added to the MS, but only with a standard storage manager.

For example, it is possible to make the FLAG column writable by creating a FLAGX column, copy FLAG to it, remove FLAG, and finally rename FLAGX to FLAG.

A.4   Lessons learned and future

Besides the gain in write speed, LofarStMan proved successful for another reason:
Accidently the data writer wrote the conjugates of the data during the first period. This could be solved elegantly by adapting LofarStMan such that version 1 conjugated the data and by setting the version to 2 for the new data writer.

MeasurementSets are often selected or sorted on TIME. This proved to be slow because the entire data file had to be traversed. Therefore a new version 3 will get an optional extra file *table.f0seqnr* containing a copy of the sequence numbers. Normally this file will be present and correspond to the data file size. If so, LofarStMan will take advantage of it. Because the file is written afterwards, it may not be present or correct in case of failures. In that case LofarStMan will get the sequence numbers from the data file.

The only real problem encountered was that an MS stored with LofarStMan could not be used in the CASA package. The reason is that CASA uses different casacore libraries than LofarStMan, so even after it has registered itself, LofarStMan was still not known to CASA. This will be solved if CASA can be built using the same libraries.

LOFAR/
USG/
Data Formats

Doc.nr.:

Rev.:

Date:

Class.:   Public

32 of 32