MS2 description for LOFAR-CS1

	Organisatie / Organization	Datum / Date
Auteur(s) / Author(s): A.P. Schoenmakers	ASTRON	01-01-2006
Controle / Checked:	ASTRON	
Goedkeuring / Approval:	ASTRON	
Autorisatie / Authorisation: **Handtekening / Signature**	ASTRON	

Plaats voor een LOGO

**LOFAR/
USG/
Data Formats**

Doc.nr.:

Rev.:

Date:

Class.: Public

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

Plaats voor een LOGO

**LOFAR/
USG/
Data Formats**

Doc.nr.:

Rev.:

Date:

Class.: Public

Table of contents:

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.:   Public

3 of 23

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/CS1 MS
This is based (and largely copied) from the original MS2 definition document. The Measurement Set v2 is described in great detail in a AIPS++ memo which can be found here:
http://aips2.nrao.edu/stable/docs/notes/229/229.html
A local copy in pdf is available here:
http://wsrtweb.nfra.nl/tmswiki/doku.php?id=wsrtinfo:wsrtmsformat

The fields given in parentheses are compulsory, and may thus be skipped. Some fields which are not relevant at all, have been skipped (see the MS2 definition document for their meaning).

2 General description of the Measurement Set

In this section I will briefly describe all tables and subtables in a Measurement set, and what type of information these contain. I will discuss individual fields in sofar as these require additional explanations to understand their exact meaning. Table items which may still require some discussion or decision, have been put in italics. These are mostly administrative fields, not the main data fields.

2.1 Structure of the MS

In the table I present a short summary of the subtables in a LOFAR-CS1 Measurement Set, and a brief explanation of their contents. In the following sections I will present each table in some more detail.

Table name	Short description of content
MAIN	Data of all samples for individual ifrs
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
HISTORY	History log of MS
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 SD)

2.2 MAIN table

The MAIN table of the MS contains the bulk of all data. For each interferometer and for each sampletime The ordering is usually time-baseline, i.e. the MAIN table is divided in subsequent time-blocks and within each timeblock 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.

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

4 of 23

2.3 ANTENNA Table

The Antenna table contains information on each (micro) station 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.

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, tyo be applied to the data in the MAIN table. I do not know if we require this for LOFAR, so I skip this table for the moment.

2.8 HISTORY table

This table allows some 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. If LOFAR will fill this table, we need to work out a convention for the fields in the table. For the moment, I'll not describe this table in detail.

2.9 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 is purely administrative info, but useful to understand the origin of the dataset.

2.10 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.

2.11 POLARIZATION table

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

2.12 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).

2.13 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/CS1 Measurement Set as it helps to clarify the goal of the observation.

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

5 of 23

2.14 SPECTRAL_WINDOW table

This table presents the frequency settings of the observation.

2.15 STATE table

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

Plaats voor een LOGO

**LOFAR/
USG/
Data Formats**

Doc.nr.:

Rev.:

Date:

Class.: Public

MAIN Table: Data, coordinates and flags					
Name	Format	Units	Measure	Default	Comments
Keywords					
MS_VERSION	Float			2.0	MS version number
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
FFED2	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				Sample number
ARRAY_ID	Int			0	Subarray 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

3   Detailed description of all MS tables

3.1   MAIN table

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 CS1, so *Ncat*=1).

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

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.:   Public

7 of 23

TIME Mid-point (not centroid) of data interval. In UTC.

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

FEED*n* Feed number (≥ 0). Points to 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 indentifier (≥ 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 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.

ARRAY_ID Subarray identifier (≥ 0), which identifies data in separate subarrays. Always 1 for CS1 (only 1 subarray defined)

OBSERVATION_ID Observation identifier (≥ 0), which identifies data from separate observations, and a direct index into the OBSERVATION sub-table *rownr*.

STATE_ID State identifier (≥ 0), and a direct index into the STATE sub-table *rownr*. Always 1 for CS1 (only one entry in STATE table)

UVW *uvw* coordinates for the baseline from ANTENNE2 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 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. The values are also in channel 0 of the DATA column.

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 CS1, the weight spectrum is determined in the correlator.

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 bad if the FLAG array element is True. This column will be altered by the Flagger program.

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-CS1 and will remain empty.

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

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

8 of 23

3.2 ANTENNA table

ANTENNA Table: Antenna Characteristics					
Name	Format	Units	Measure	Default	Comments
Data					
NAME	String			LOFAR	Unique antenna names
STATION	String				Station name
TYPE	String			GROUND-BASED	Antenna Type
MOUNT	String			FIXED	Antenna Mounting
POSITION	Double (3)	m	POSITION		Antenna X,Y,Z phase reference positions
OFFSET	Double (3)	m	POSITION	0	Axes offset of mount to FEED REFERENCE point
DISH_DIAMETER	Double	m		150	Approx. Station size
Flag information					
FLAG_ROW	Bool			False	Row flag

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

NAME Antenna name; This will be "LOFAR".
STATION Station name. Stations will be named as follows: CSxxx for Core Stations (xxx is a 3-digit number; the station numbers will be equal to the physical numbering); RSxxx for Remote Stations; YYxxx for foreign station; YY will be the country code in usenet/internet format). For substations (only relevant to the CS1-phase) we will use the designation: CSxxx.y where y is a 1-digit substation number that equals the number of the input line to the station's RSP board (0-7).
TYPE Antenna type. Reserved keywords include: ("GROUND-BASED" - conventional antennas; "SPACE-BASED" - orbiting antennas; "TRACKING-STN" - tracking stations). We will use "GROUND-BASED".
MOUNT 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.). We will use "FIXED".
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. The reference point is the point on the az or ha axis closest to the el or dec axis.
OFFSET Axes offset of mount to feed reference point.
DISH_DIAMETER Nominal diameter of dish, as opposed to the effective diameter. For LOFAR this will be the approximate size of a (sub)station. For now, this is set at 150 meter.
FLAG_ROW 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.

Time variant LOFAR 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., reference position). There are two options when this becomes relevant:
- A new entry is made in the Antenna table when a change occurs in the reference position of an station. In the MAIN table there must be a reference to the new entry. However, this makes the administration rather cumbersome, as the Antenna table itself has no time-related column containing the time span when a particular entry is valid. Alternatively, we could add such a column ourselves, but that would make the MS somewhat non-standard.

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

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

9 of 23

- Time variance of ANTENNA properties may also be handled in the FEED table (see below). In this case, a new entry is made in the FEED table when a station characteristic changes. A reference position change can be mimicked as a change in the FEED position.

In both cases, only very particular changes in the station characteristics (mainly reference position) can be handled anyway. For CS1 these discussions are not relevant yet, but they will become relevant once LOFAR will grow beyond CS1.

3.3 DATA_DESCRIPTION Table

DATA_DESCRIPTION Table: Frequency/polarization characteristics					
Name	Format	Units	Measure	Default	Comments
Data					
SPECTRAL_WINDOW_ID	Int				
POLARIZATION_ID	Int				
Flag information					
FLAG_ROW	Bool				Row flag

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

SPECTRAL_WINDOW_ID Spectral window identifier; direct index into the SPECTRAL_WINDOW sub-table.
POLARIZATION_ID Polarization identifier (≥ 0); direct index into the POLARIZATION sub-table.
FLAG ROW True if the row does not contain valid data; does not imply flagging in MAIN.

3.4 FEED table

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

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

10 of 23

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	Nr. 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	Position of feed relative to feed reference position for this antenna
RECEPTOR ANGLE	Double (NUM RECEPTORS)	rad			The reference angle for polarization.

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.

ANTENNA_ID Antenna 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 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.
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 center of the beam for this feed. Expressed in a linearly polarized basis (*ex,ey*) using the IEEE convention. For CS1, this will be (1,0),(0,1)
POSITION Offset of feed relative to the feed reference position for this antenna (see ANTENNA sub-table). We'll assume this is (0,0,0)
RECEPTOR_ANGLE Polarization reference angle. Converts into parallactic angle in the sky domain.

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

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

11 of 23

3.5 FIELD table

FIELD Table: Field positions for each source					
Name	Format	Units	Measure	Default	Comments
Data					
NAME	String				Name of field
CODE	String				Special characteristics
TIME	Double	s	EPOCH		Time origin for the directions and rates
NUM_POLY	Int			0	Polynomial series order
DELAY_DIR	Double (2, NUM_POLY+1)	rad	DIRECTION		Direction of delay center.
PHASE_DIR	Double (2, NUM_POLY+1)	rad	DIRECTION		Direction of Phase center
REFERENCE_DIR	Double (2, NUM_POLY+1)	rad	DIRECTION		Direction of Reference center
SOURCE_ID	Int			-1	Index in SOURCE table
Flag information					
FLAG_ROW	Bool			False	Row flag

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

NAME Field name; user specified.
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 center; 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) center; 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 center; 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 center, 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.
FLAG_ROW True if data in this row are invalid, else False. Does not imply flagging in MAIN.

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

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

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

12 of 23

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			LofarSim	Name of observer
LOG	String (*)				Observing log
SCHEDULE_TYPE	String			LOFAR	
SCHEDULE	String (*)			[corrSchedule]	Project Schedule
PROJECT	String			LofarSim	Project identification
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.

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").
SCHEDULE Unmodified schedule file, of the type specified, and as used by the instrument.
PROJECT Project code (e.g. "BD46")
RELEASE_DATE Project release date. This is the date on which the data may become public.
FLAG_ROW Row flag. True if data in this row is invalid, but does not imply any flagging in MAIN.

TBD: What values should be entered in this table?

3.7 POINTING table

POINTING Table: Antenna Pointing information					
Name	Format	Units	Measure	Default	Comments
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

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

13 of 23

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 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 starttime 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, equal this 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 labeling 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,YY,XY,YX) 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,YY,XY,YX) 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 is the data in this row are not valid, but does not imply the flagging of any DATA in MAIN.

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

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

14 of 23

3.9 PROCESSOR table

PROCESSOR Table: Processor (backend) information					
Name	Format	Units	Measure	Default	Comments
Data					
TYPE	String			CORRELATOR	
SUB_TYPE	String			CEP	
TYPE_ID	Int			0	
MODE_ID	Int			0	
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-CS1 this will be 'CORRELATOR'
SUB_TYPE Processor sub-type, e.g. "GBT" or "JIVE". We'll use 'CEP'.
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. [TBD: has this meaning for CEP?]
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 sub-table). [TBD: Are there different modes for CEP?].
FLAG_ROW Row flag. True if data in the row is not valid, but does not imply flagging in MAIN.

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				Nr. 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.

Plaats voor een LOGO	**LOFAR/ USG/ Data Formats**	**Doc.nr.:**
		Rev.:
		Date:
		Class.: Public

TIME Mid-point of the time interval for which the data in this row is valid. Required to use the same TIME Measure reference as in MAIN.
INTERVAL Time interval for which the data in this row is 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.

TBD: Will LOFAR use the calibration group/code in its pipeline? How to deal with multi-beam observations?

SPECTRAL WINDOW Table: Frequency setup information					
Name	Format	Units	Measure	Default	Comments
Data description					
NUM_CHAN	Int			256	Nr. 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		Center 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				The IF conversion chain
FREQ_GROUP	Int				
FREQ_GROUP_NAME	String				
Flag information					
FLAG_ROW	Bool			False	Row flag

3.11 SPECTRAL_WINDOW table

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

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

16 of 23

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 is 256, always.

NAME Spectral window name; for CS1 we will use SB-x where x is a number starting from 0 that indicates the subband number used. A LOFAR-CS1 MS may have multiple subbands, but a subband can only be in a single MS.

REF_FREQUENCY The reference frequency. A 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.

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). [**TBD**: Relevance to LOFAR-CS!]

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, calibration is done per subband, but all subbands must be in a main band. We can therefore identify the main band here. There a four main bands: 30 – 80 (Low band), 110 – 190 , 170 – 230, 210 – 270 MHz . This can be numbered as 1,2,3,4 respectively.

FREQ_GROUP_NAME The frequency group name. For LOFAR we can use the name of one of the four main bands (LB, HB1, HB2, HB3)

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

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

17 of 23

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-CS1. 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 in how far this is actually useful needs to be discussed.

3.13 Schema of the LOFAR-CS1 MS

The figure below shows a graphical representation of the LOFAR-CS1 Measurement Set. All Subtables and columns in the MS are presented.
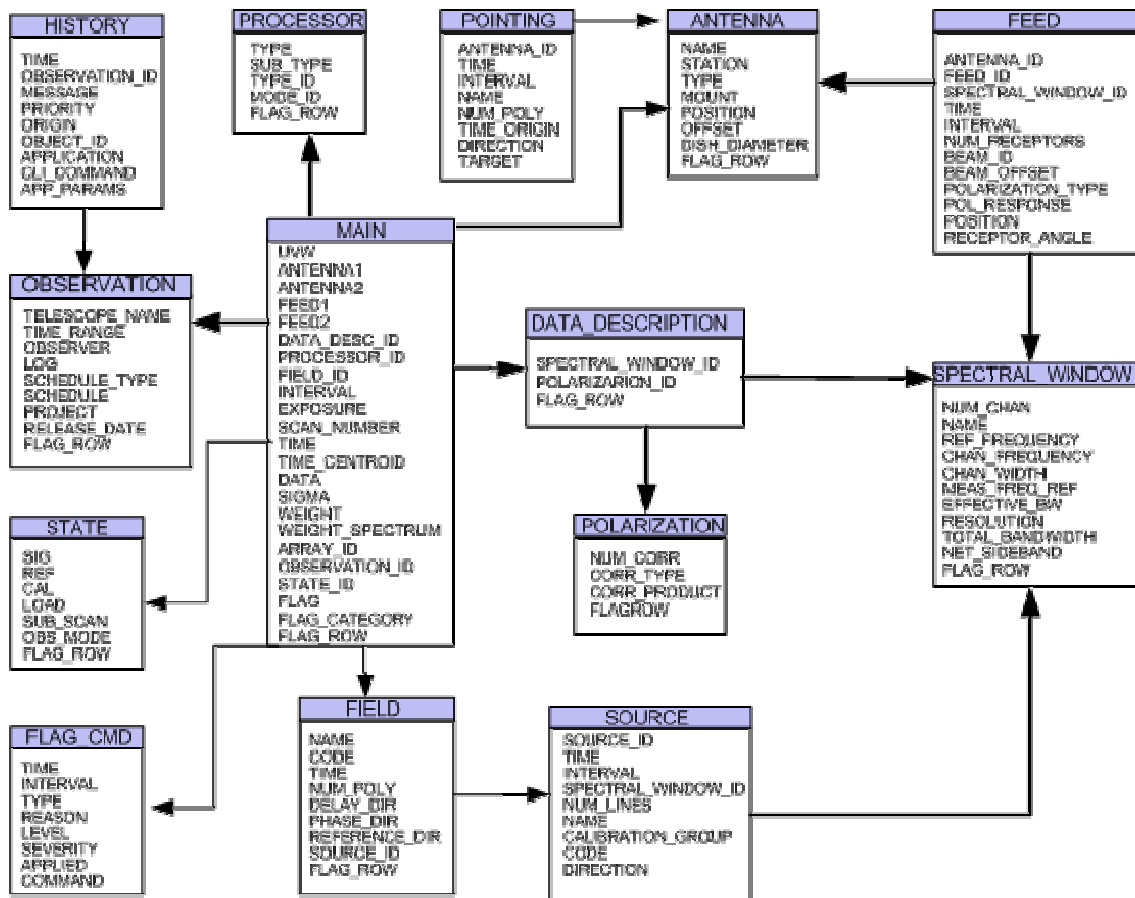
Plaats voor een LOGO

**LOFAR/
USG/
Data Formats**

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

18 of 23

Figure 1: Schema of the LOFAR-CS1 Measurement Set

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.: Public

19 of 23

A Appendix A: Translating the parset file.

The storage process that writes the MS obtains most of its meta-data through parset files. In this Appendix I will present how the translations are being done, and which Measurement Set fields are filled with which parset parameter values.

The Parset files are ASCII files containing lines of the type 'keyword = value(s)'. These are the basic command files for all LOFAR processes. Eventually the parset files are produced for each observation directly from SAS, but currently this is still a manual process.

A typical parset file contains many lines. The majority is used to setup the connections between the various hardware pieces (e.g., through MAC or IP addresses). Some of the parameters are related to the observation to be conducted. An example is given below (from the parset file L2006_00417.MS.parset):

```
Observation.BeamDirections = [0, 0.92]
Observation.DirectionType = J2000
Observation.NBeams = 1
Observation.NChannels = 256
Observation.NPolarisations = 2
Observation.NStations = 16
Observation.NSubbandSamples = 155648
Observation.NSubbands = 1
Observation.PositionType = ITRF
Observation.RefFreqs = [60000000]
Observation.SampleRate = 156250
Observation.StartTime = 1164387436
Observation.StationPositions = [0.119879729, 0.92350930899999994, -
472.88506899999999, 0.119887278, 0.92351232000000005, -472.88491699999997,
0.119878255, 0.92350892299999998, -472.88505099999998, 0.119884052,
0.92351078600000003, -472.885062, 0.119851103, 0.92348175600000004, -
472.88248299999998, 0.11985772, 0.92348559600000002, -472.882901,
0.119864071, 0.92348159500000004, -472.882206, 0.119857454,
0.92347775600000004, -472.88176600000003, 0.1198682580000001,
0.92353613199999995, -472.88706000000002, 0.1198748760000001,
0.92353997200000004, -472.88681000000003, 0.11989796, 0.92353199900000005, -
472.88746800000001, 0.119891342, 0.92352816000000004, -472.88675499999999,
0.1199584280000001, 0.92350928600000004, -472.88546500000001,
0.1199650460000001, 0.92351312500000005, -472.88584800000001,
0.1199713969999999, 0.92350912500000004, -472.88518699999997,
0.1199647789999999, 0.92350528499999995, -472.88442500000002]
Observation.StopTime = 1164402434
Storage.IntegrationTime = 60
Storage.MSName = /data/L2006_00417.MS
Storage.StorageStationNames = ['CS10_dipole0', 'CS10_dipole4',
'CS10_dipole8', 'CS10_dipole12', 'CS01_dipole0', 'CS01_dipole4',
'CS01_dipole8', 'CS01_dipole12','CS08_dipole0', 'CS08_dipole4',
'CS08_dipole8', 'CS08_dipole12', 'CS16_dipole0', 'CS16_dipole4',
'CS16_dipole8', 'CS16_dipole12']
```

How and where are these fields entered in the MS? Let's pick out the associated lines of the parset file.

A.1   Pointing information

```
Observation.BeamDirections = [0, 0.92]
Observation.DirectionType = J2000
```

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

Doc.nr.:

Rev.:

Date:

Class.:   Public

20 of 23

```
Observation.NBeams = 1
```

These fields give the RA, Dec coordinates of the beam in rad. They are entered in the following MS tables and fields.

- Table *Pointing*; Columns *Direction, Target*; `Observation.BeamDirections`. The `Observation.BeamDirections` field goes into the columnheader as keyword.
- Table *Field*; Columns *Delay_Dir, Phase_Dir, Reference_Dir*. The `Observation.DirectionType` field goes into the columnheader as keyword.
- Table *Source*; Column *Direction*; `Observation.BeamDirections`. The `Observation.DirectionType` field goes into the columnheader

As there is no source/target name supplied in the parset file, these are defaulted to "Beam-0". This is entered in the *Field::Name* field.

A.2 Antenna information

Details of the Antennae used are observation dependent . The relevant parset lines are:

```
Observation.PositionType = ITRF
Observation.NStations = 16
Observation.StationPositions = [0.119879729, … etc…]
Storage.StorageStationNames = ['CS10_dipole0', etc…]
```

This information is fed into the ANTENNA table, in the following columns:
Name: `Storage.StorageStationNames`
Position: `Observation.StationPositions` (preferably in X,Y,Z (m) ITRF; this comes from `Observation.PositionType` and must be fed into the *Position* column header keywords).

A.3 Frequency information

The frequency setup is handed over through the parset fields:

```
Observation.SampleRate = 156250
Observation.NChannels = 256
Observation.NSubbands = 1
Observation.RefFreqs = [60000000]
```

This is translated into the *SPECTRAL_WINDOW* table as follows:
The number of rows in the *SPECTRAL_WINDOW* table is given by the `Observation.NSubbands`; this is only for LOFAR-CS1 at this point. Later, the data will be distributed over the 12 storage nodes with each node handling a part of the total number of subbands.
The `Observation.RefFreqs` field gives the reference frequency, or the midpoint frequency of the observation for each subband.
The total bandwidth of the subband is given by `Observation.SampleRate`. This can either be 156250, or ???. This is in Hz.
From these numbers all information in the SPECTRAL_WINDOW table is determined:

Num_chan: `Observation.NChannels`.
Total_bandwidth: `Observation.SampleRate` (in Hz).
Ref_frequency: `Observation.RefFreqs` (this is in Hz).
Chan_width: Array of `Observation.NChannels` values which contains `Observation.SampleRate` divided by `Observation.NChannels` (in Hz).

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

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

21 of 23

Effective_bw and *Resolution* are equal to *Chan_width*.
Chan_freq: Array of `Observation.NChannels` values. Startvalue is *Ref_frequency* minus *Total_bandwidth* / 2 + (*Total_bandwidth* / *Num_chan*) / 2. It is assumed that the *Ref_frequency* is the value at the bottom side of channel `Observation.NChannels/2` (first channel is channel 0).
Name contains the string SB-x, where x is the subband number.

A.4 Time issues

Samples are correlated at a fundamental rate (determined by the samplerate at the stations). Before storing the data in the MS, a number of samples are integrated and averaged to a given integration time. The integration time (or rather, the number of samples to be integrated before writing them into the MS as a single sample) is user-defined in the parset file. Startime and endtime of the observation are also user defined and handed over in the parset file, and used at several locations in the MS. The parset lines related to time keeping are:

```
Observation.NSubbandSamples = 155648
Observation.SampleRate = 156250
Observation.StartTime = 1164387436
Observation.StopTime = 1164402434
Storage.IntegrationTime = 60
```

The `Observation.StartTime` and `Observation.StopTime` are given in Unix seconds, i.e. seconds since Jan 1, 1970 in UT.
The fundamental correlator integration time is given by dividing `Observation.NSubbandSamples` and `Observation.SampleRate`, which for this example yields 0.9961472 seconds. The number given by `Observation.NSubbandSamples` can be determined as follows:
`256 * (int)(Observation.SampleRate/(16*256))`
This is related to the polyphase filtering used and the need to work with integer numbers.
The number of samples integrated into a single row in the MAIN table of the MS is given by `Storage.IntegrationTime`. The total sample integration time in this example is thus: 59.768832 seconds. This is also the separation time between consecutive samples.
Based on these numbers the following fields are filled:
Main table: *Exposure* and *Interval* (set equal): total sample integration time
Time and *Time_centroid* (which are set equal): This is Starttime + *Scan_number* * *Interval*; this is also used to calculate the *UVW* coordinates of the baseline.
Observation table: *Timerange*: this is `Observation.StartTime, Observation.StopTime`
Pointing table: Time and Interval: Observation midtime and duration from `Observation.StartTime, Observation.StopTime`.
Source table: *Time* and *Interval*: as in *Pointing* table.
Feed table: *Time* and *Interval*: as in *Pointing* table.
Field table: *Time*: `Observation.StartTime`.

A.5 Polarization

Polarization information is stored at two locations in the MS: The Polarization subtable and the Feed subtable. Indirectly, the number of polarizations used and correlated also determines the dimension of the fields that stores the data and the data-related information.
The parset file only has one field for polarization:

```
Observation.NPolarisations = 2
```

For CS1, the polarization is always linear, thus the polarizations used are (X,Y). This is put in the *Feed* subtable:
Polarization_type: Fixed
Polarization_respons: Fixed

Plaats voor een LOGO

LOFAR/ USG/ Data Formats

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

22 of 23

Receptor_angle: Fixed.
Num_receptors: Fixed; equals `Observation.NPolarisations`

The BGL correlator calculates all cross-products and the data in the MS thus contains all of these: XX, XY, YX, YY.
This information is stored in the *Polarization* subtable:
Corr_type: contains the AIPS++-internal code for the four cross products.
Corr_product: contains the matrix that determines the four cross-products from the vector (X,Y).
Num_corr: Gives the number of cross-products (i.e., 4).

A.6 DataSet naming

The name with which the MS is written to disk is taken directly from the parset file:

```
Storage.MSName = /data/L2006_00417/SBx.MS
```

Note, that in case more than one storage node is used in writing the data (e.g, when multiple subbands are written), this parameter is an array holding the MS names to be used for each storage node.

Plaats voor een LOGO

**LOFAR/
USG/
Data Formats**

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

23 of 23