A brief introduction to the Measurement Set and the ALMA Science Data Model

Dirk Petry (ESO), January 2011

Outline:

→ **Radio Data** — from the point of view of a database

→ **The Measurement Set** — the ALMA analysis data format

→ **The ASDM** — the ALMA archive format and its conversion to MS

→ **For the ALMA Cycle 0 PI** — MS examples

→ **Summary**

- **Radio Astronomical Observatories: single-dish or interferometer**
- **Single dish data are *real numbers (Amplitudes)***
- **Interferometer data are *complex numbers (= 2 real numbers: Amplitude and Phase), the so-called Visibilities***
- **There is one Amplitude (single-dish) or Visibility (interferometer) for each *Spectral Channel* and for each *Polarization* (single-dish) or *Correlation* (interferometer)**
- **Recorded in *Integrations with Timestamps*: one for each *Antenna* (single-dish) or antenna-pair a.k.a. *Baseline* (interferometer)**
- **A set of consecutive integrations observing the same celestial direction (*Field*) is called a *Scan***
- **A set of one or more Scans is one *Observation***
- **In addition to the bulk data (Amplitudes/Visibilities), the data analysis requires meta-data concerning:**
  - **-- the Antennas: positions, diameters, ...**
  - **-- the Receivers/Feeds on the Antennas: position, sensitivity, ...**
  - **-- the Spectral Channels: frequencies, noise, ... (the *Spectral Window* setup)**
- **Not always required but stored with the bulk data for completeness are**
  - **-- Sources (known celestial objects) in the field of view, Spectral Lines, ...**
  - **-- the Weather: Temperatures, Water Vapour in the Atmosphere, ...**
  - **-- the Observing Log, Data from the online Calibration, and more**

Radio Data

single-dish

Amplitudes

(obtained from total power or spectrometer measurements, or from autocorrelation)

Antenna A12, Timestamp 2:49:45 h

Amplitudes	Spectral Channel				
	0	*1*	*2*	*3 ...*
Polarization					
X	*0.1*	*2.3*	*2.1*	*1.9*
Y	*1.2*	*3.3*	*8.9*	*6.5*
.	*...*				

interferometer

Visibilities (Amp,Phase)

(obtained from a correlator)

Baseline (A3,A12), Timestamp 2:49:45 h

Visibilities	Spectral Channel				
	0	*1*	*2*	*3 ...*
Correlation					
XX	*(0.2,2.1)*	*(3.2,2.0)*	*(3.0,1.9)*	*(2.8,1.8) ...*
YY	*(1.3,1.9)*	*(2.8,2.0)*	*(7.6,1.8)*	*(5.5,1.9) ...*
.	*...*				

In addition, for each Amp./Vis. have a "Flag" == marker for bad data

Flags	Spectral Channel				
	0	*1*	*2*	*3 ...*
Polarization					
X	*0*	*0*	*1*	*0*
Y	*1*	*1*	*1*	*1*
.	*...*				

Flags	Spectral Channel				
	0	*1*	*2*	*3 ...*
Correlation					
XX	*0*	*0*	*1*	*0 ...*
YY	*1*	*1*	*1*	*1 ...*
.	*...*				

The Measurement Set

The *Measurement Set* (MS)

- **developed by Cornwell, Kemball, & Wieringa between 1996 and 2000**

- **designed to store both interferometry and single-dish data**

- **supports (in principle) any setup of radio telescopes**

- **supports calibration of the data via the Measurement Equation**

- *relational database system with fixed structure*

- *consists of a main table with 12 required sub-tables + several optional ones*

- *MS = table for visibilities or amplitudes + auxiliary sub-tables*

- *MS v2.0* **described in http://casa.nrao.edu/Memos/229.html**

The Measurement Set

The MS in *CASA*

- in *CASA* implemented as *CASA Tables*

- uses OS directory structure (need to copy with `cp -R`, remove with `rm -r`)

- visibilities stored in the MAIN table

- manipulate an MS with the `ms` and the `tb` tool or with `browsetable()`

- during processing, CASA may add "scratch columns" to the MS main table

The Measurement Set

The MS in *CASA*

- during processing, CASA may add "scratch columns" to the MS main table

    DATA   **- the original visibilities from when the MS was created**
                           (not a scratch column, always present)

    CORRECTED_DATA   **- calibrated visibilities obtained from DATA**
                           (initialised with a copy of DATA)

    MODEL_DATA   **- Fourier transform of a model image**
                           (initialised to 1)

CASA Measurement Set

Example of a CASA *Measurement Set in the file system* on disk

```
$ tree AT352_A071103-K

AT352_A071103-K
|-- ANTENNA                |-- FIELD               |-- POINTING            |-- SPECTRAL_WINDOW |-- table.f1_TSM1
|   |-- table.dat          |   |-- table.dat       |   |-- table.dat       |   |-- table.dat    |-- table.f2
|   |-- table.f0           |   |-- table.f0        |   |-- table.f0        |   |-- table.f0     |-- table.f2_TSM1
|   |-- table.info         |   |-- table.f0i       |   |-- table.f0i       |   |-- table.f0i    |-- table.f3
|   `-- table.lock         |   |-- table.info      |   |-- table.f1        |   |-- table.info   |-- table.f3_TSM1
|-- DATA_DESCRIPTION       |   `-- table.lock      |   |-- table.info      |   `-- table.lock   |-- table.f4
|   |-- table.dat          |-- FLAG_CMD            |   `-- table.lock      |-- STATE           |-- table.f4_TSM1
|   |-- table.f0           |   |-- table.dat       |-- POLARIZATION        |   |-- table.dat    |-- table.f5
|   |-- table.info         |   |-- table.f0        |   |-- table.dat       |   |-- table.f0     |-- table.f5_TSM1
|   `-- table.lock         |   |-- table.info      |   |-- table.f0        |   |-- table.info   |-- table.f6
|-- DOPPLER                |   `-- table.lock      |   |-- table.f0i       |   `-- table.lock   |-- table.f6_TSM1
|   |-- table.dat          |-- HISTORY             |   |-- table.info      |-- table.dat       |-- table.f7
|   |-- table.f0           |   |-- table.dat       |   `-- table.lock      |-- table.f0        |-- table.f7_TSM1
|   |-- table.info         |   |-- table.f0        |-- PROCESSOR           |-- table.f1        |-- table.f8
|   `-- table.lock         |   |-- table.info      |   |-- table.dat       |-- table.f10       |-- table.f8_TSM1
|-- FEED                   |   `-- table.lock      |   |-- table.f0                            |-- table.f9
|   |-- table.dat          |-- OBSERVATION         |   |-- table.info                          |-- table.f9_TSM0
|   |-- table.f0           |   |-- table.dat       |   `-- table.lock                          |-- table.info
|   |-- table.f0i          |   |-- table.f0        |-- SORTED_TABLE                            `-- table.lock
|   |-- table.info         |   |-- table.info      |   |-- table.dat
|   `-- table.lock         |   `-- table.lock      |   `-- table.info            15 directories, 88 files
```

CASA design and implementation
The Measurement Set

MAIN   none   -   -   ANTENNA_ID / FEED_ID / DATA_DESC_ID / PROCESSOR_ID / (PHASE_ID) / FIELD_ID / (PULSAR_GATE_ID) / ARRAY_ID / OBSERVATION_ID / STATE_ID

ANTENNA   ANTENNA_ID   row number   MAIN / FEED / FREQ_OFFSET / POINTING / SYSCAL / WEATHER   (ORBIT_ID) / (PHASED_ARRAY_ID)

POLARIZATION   POLARIZATION_ID   row number   DATA_DESCRIPTION   none

FEED   FEED_ID   explicit   MAIN / FREQ_OFFSET / SYSCAL   ANTENNA_ID / SPW_ID / BEAM_ID / (PHASED_FEED_ID)

(SOURCE)   SOURCE_ID   explicit   (DOPPLER) / FIELD   SPW_ID / (PULSAR_ID)

(FREQ_OFFSET)   none   -   -   ANTENNA_ID / FEED_ID / SPW_ID

DATA_DESCRIPTION   DATA_DESC_ID   row number   MAIN   SPW_ID / POLARIZATION_ID / (LAG_ID)

(DOPPLER)   DOPPLER_ID   explicit   SPW_ID   SOURCE_ID / TRANSITION_ID

(SYSCAL)   none   -   -   ANTENNA_ID / FEED_ID / SPW_ID

PROCESSOR   PROCESSOR_ID   row number   MAIN   TYPE_ID / MODE_ID / (PASS_ID)

POINTING   none   -   -   ANTENNA_ID / POINTING_MODEL_ID

FIELD   FIELD_ID   row number   MAIN   SOURCE_ID / (EPHEMERIS_ID)

OBSERVATION   OBSERVATION_ID   row number   MAIN / HISTORY   none

(WEATHER)   none   -   -   ANTENNA_ID

STATE   STATE_ID   row number   MAIN   none

HISTORY   none   -   -   OBSERVATION_ID / OBJECT_ID

SPW   SPW_ID   row number   DATA_DESCRIPTION / FEED / FREQ_OFFSET / SOURCE / SYSCAL   (RECEIVER_ID) / (DOPPLER_ID) / (ASSOC_SPW_ID)

FLAG_CMD   none   -   -   none

Legend:

[Table Name]   [Key defined in this table]   [key definition method]   [referenced by]   [referenced keys] (optional) / reference to table outside the MS definition

Level 1: Tables not referenced by other tables

Level 2: Tables referenced by level 1

Level 3: Tables referenced by level 2

V1, D.Petry, 13.2.09

CASA internal and external visibility data formats

- **Internal CASA visibility data format is the Measurement Set (MS)**

- **Presently supported input formats:**

  ALMA: ALMA Science Data Model (ASDM) - importasdm

  **EVLA: Science Data Model (SDM, essentially the same as the ASDM)
        - importevla**

  VLA: VLA archive format - importvla

  EVN, eMERLIN et al.: FITS-IDI - importfitsidi

  and the general transport format uvfits - importuvfits

The ASDM

Facts about the ASDM (ALMA) and SDM (EVLA)

- **relational database system with fixed structure**

- **version 1 consists of set of up to 56 tables (64 in latest version)**

- **contains much more setup and calibration information than the MS**

- **uses OS directory structure (need to copy with `cp -R`, remove with `rm -r`)**

- **visibilities stored in the MAIN table**

- **on disk, all table descriptions in XML, some table data in binary MIME format files**

- **import into CASA using the task `importasdm` or `importevla`**

- **since release 3.1 there is also `exportasdm` (MS to ASDM)**

The ASDM Tables (model version 1.61)

Visibilties
- MainTable (+binary)

Observation Layout (→APDM)
- ExecBlockTable
- SBSummaryTable

Setup/History
- AlmaRadiometerTable
- AnnotationTable
- AntennaTable
- BeamTable
- ConfigDescriptionTable
- CorrelatorModeTable
- DataDescriptionTable
- DelayModelTable
- DopplerTable
- EphemerisTable
- FeedTable
- FieldTable
- FlagCmdTable*
- FocusModelTable
- FocusTable
- FreqOffsetTable
- GainTrackingTable

Setup/History (ctnd.)
- HistoryTable*
- HolographyTable
- ObservationTable*
- PointingModelTable
- PointingTable (+binary)
- PolarizationTable
- ProcessorTable
- ReceiverTable
- ScaleTable*
- ScanTable
- SeeingTable*
- SourceTable
- SpectralWindowTable
- SquareLawDetectorTable
- StateTable
- StationTable
- SubscanTable
- SwitchCycleTable
- SysCalTable
- SysPowerTable*
- TotalPowerTable*
- WeatherTable
- WVMCalTable

Calibration
- CalAmpliTable
- CalAtmosphereTable
- CalBandpassTable
- CalCurveTable
- CalDataTable
- CalDelayTable
- CalDeviceTable
- CalFluxTable
- CalFocusModelTable
- CalFocusTable
- CalGainTable (not used)
- CalHolographyTable
- CalPhaseTable
- CalPointingModelTable
- CalPointingTable
- CalPositionTable
- CalPrimaryBeamTable
- CalReductionTable
- CalSeeingTable
- CalWVRTable

Total
- **63 Tables** (* 7 not in version 1.0)

The ASDM Tables

- Tables needed for a minimal ASDM:

Main

ConfigDescription

Field

Station

Antenna

State

ExecBlock

Feed

DataDescription

Processor

SwitchCycle

Station

SBSummary

Receiver

SpectralWindow

Polarization

CorrelatorMode

Scan

SubScan

The ASDM - example of an ASDM on disk

```
uid___X02_X56142_X1
|-- ASDM.xml                        |-- AlmaRadiometer.xml        |-- PointingModel.xml
|-- ASDMBinary                      |-- Antenna.xml               |-- Polarization.xml
|     |-- uid___X02_X5614c_X1       |-- CalAmpli.xml              |-- Processor.xml
|     |-- uid___X02_X5614c_X2       |-- CalCurve.xml              |-- Receiver.xml
|     |-- uid___X02_X5614c_X3       |-- CalData.xml               |-- SBSummary.xml
|     |-- uid___X02_X56151_X1       |-- CalDevice.xml             |-- Scan.xml
|     |-- uid___X02_X56151_X2       |-- CalPhase.xml              |-- Source.xml
|     |-- uid___X02_X56151_X3       |-- CalPointing.xml           |-- SpectralWindow.xml
|     |-- uid___X02_X5615a_X1       |-- CalReduction.xml          |-- State.xml
|     |-- uid___X02_X5615a_X2       |-- CalSeeing.xml             |-- Station.xml
|     |-- uid___X02_X5615a_X3       |-- CalWVR.xml                |-- Subscan.xml
|     |-- uid___X02_X56160_X1       |-- ConfigDescription.xml     |-- SwitchCycle.xml
|     |-- uid___X02_X56160_X2       |-- CorrelatorMode.xml        `-- Weather.xml
                                    |-- DataDescription.xml
...                                 |-- ExecBlock.xml             1 directory, 251 files
                                    |-- Feed.xml
|     |-- uid___X02_X563c3_X2       |-- Field.xml
|     |-- uid___X02_X563c3_X3       |-- Focus.xml
|     |-- uid___X02_X563cc_X1       |-- FocusModel.xml
|     |-- uid___X02_X563cc_X2       |-- Main.xml
|     `-- uid___X02_X563cc_X3       |-- Pointing.bin
```

The ASDM to MS conversion

CASA works on MSs. Conversion of ASDMs to MSs done via the CASA task

importasdm

```
#  importasdm :: Convert an ALMA Science Data Model observation into a CASA visibility file
asdm                  = ''    # Name of input asdm directory (on disk)
vis                   = ''    # Root name of the ms to be created.
singledish            = False # Set true to output single-dish data format
corr_mode             = 'all' # correlation mode to be considered on input.
srt                   = 'all' # spectral resolution type: fr, ca, bw, or all
time_sampling         = 'all' # time sampling: i, si, or all
ocorr_mode            = 'ca'  # output data for correlation mode AUTO_ONLY (ao)
                              # or CROSS_ONLY (co) or CROSS_AND_AUTO (ca)
compression           = False # Flag for turning on data compression
asis                  =  ''   # Creates verbatim copies of the ASDMtables in the ouput MS
                              # table names separated by spaces; A * wildcard is allowed.
wvr_corrected_data    = 'no'  # SDM binary data to fill the DATA column: no, yes, both (2 MSs)
scans                 =  ''   # semicolon separated list of scans, e.g."0:1,1:2~6,8;2:,3:24~30"
ignore_time           = False # Process all rows of Feed, History, Pointing, Source, SysCal,
                              # CalDevice, SysPower, Weather independently of scan selection
process_syspower      = True  # Process SysPower table
process_caldevice     = True  # Process CalDevice table
process_pointing      = True  # Process Pointing table
verbose               = False # Diagnostic output during filling
overwrite             = False # Overwrite an existing MS
showversion           = False # Report the version of asdm2MS being used
useversion            = 'v3'  # Selects the version of asdm2MS to be used: v2 or v3
async                 = False # If true the taskname must be started using importasdm(...)
```

For the ALMA Cycle 0 PI

A typical ALMA Cycle 0 MS

```
2011.0.00101.S/sg_ouss_id/group_ouss_id/member_ouss_id/raw/uid___A002_X2f146f_X6f.ms.split
```

ANTENNA	OBSERVATION	SYSPOWER	table.f16	table.f21_TSM0	table.f7
ASDM_ANTENNA	POINTING	table.dat	table.f17	table.f22	table.f8
ASDM_STATION	POLARIZATION	table.f1	table.f17_TSM1	table.f22_TSM1	table.f9
CALDEVICE	PROCESSOR	table.f10	table.f18	table.f23	table.info
DATA_DESCRIPTION	SORTED_TABLE	table.f11	table.f18_TSM1	table.f23_TSM1	table.lock
FEED	SOURCE	table.f12	table.f19	table.f3	WEATHER
FIELD	SPECTRAL_WINDOW	table.f13	table.f2	table.f4	
FLAG_CMD	STATE	table.f14	table.f20	table.f5	
HISTORY	SYSCAL	table.f15	table.f21	table.f6	

Summary

- **One modern way (and the ALMA way) of storing radio data for analysis is the MS**
- **Partially derived from the MS, but much extended, is the ALMA archive format, the ASDM**
- **developed mostly by Francois Viallefond, Robert Lucas, and Michel Caillat**
- **uses XML and MIME Binary format**
- **is essentially a relational database, each table is an independent archive object and the whole ASDM is one as well**
- **in order to analyse ALMA data, the ASDMs from the archive need to be translated to an MS (*importasdm* task in CASA*)**
- ***in cycle 0, the data are delivered as MSs for simplicity***
- **lastest MS version is 2.0**
- **latest ASDM major version is 1.0, latest minor version is 1.61**
- **the MS is essentially frozen, the ASDM is still being refined**