Scantable vs. MS

Takeshi Nakazato (NAOJ)

Contents

- Motivation

- Scantable: Merits and Demerits

- MS: Merits and Demerits

- Candidate Plans
  - Plan 1: Extend Scantable Definition
  - Plan 2: Switch to MS

- Summary

Motivation

- There is an information loss in MS to Scantable conversion
  - Interferometry specific information is acceptable
  - Common information is critical
  - Single-dish specific information is critical
  - cf. CASA requirements
- There are some uncertainties in Scantable definition
  - Especially for IF and polarization
  - The PKSMS2writer sometimes fails
- In view of CASA user and/or developer
  - Coexistence of two data format in one software
  - Same development for two data format (e.g. plotter)

Which Information Lost?

- MS -> Scantable
  - SCAN_NUMBER
    - SCANNO renumbered!
  - ANTENNA subtable
    - Only NAME and POSITION
    - DISH_DIAMETER lost! (hard coded)
  - FEED subtable
    - Only ID (BEAMNO)
    - POLARIZATION_TYPE lost! (POLTYPE, but hard coded)
  - FLAG_CMD subtable
    - completely lost
  - FIELD subtable
    - Only NAME is (FIELDNAME)
  - HISTORY subtable
    - completely lost

Which Information Lost?

- MS -> Scantable
  - OBSERVATION subtable
    - Only OBSERVER and PROJECT in first row
    - TELESCOPE_NAME lost! (cf. NAME in ANTENNA table)
    - SCHEDULE lost!
  - POINTING subtable
    - Truncated! (may be critical for OTF observation)
  - POLARIZATION subtable
    - Only maximum NUM_CORR
    - CORR_TYPE, CORR_PRODUCT lost! (hard coded)
  - PROCESSOR subtable
    - completely lost
  - SOURCE subtable
    - No problem

Which Information Lost?

- MS -> Scantable
  - SPECTRAL_WINDOW subtable
    - Different type of format (effective amount of information is reduced?)
    - ASSOC_SPW_ID lost!
  - STATE subtable
    - Only SIG, REF, and OBS_MODE (not completely)
    - CAL, LOAD lost!
    - SUB_SCAN lost! (similar to CYCLENO, but different)
  - SYSCAL subtable
    - Only TCAL and TSYS
    - TRX, TANT, TSKY lost!
  - WEATHER subtable
    - IONOS_ELECTRON, DEW_POINT, H2O lost!

Requirements: Data Format

- *OL-3.1-R2 The Package shall be able to handle the integrated data objects corresponding to the observational programs carried out by ALMA. These objects may be implemented in any manner appropriate, though relations between the components of the object must be maintained through some mechanism. These include:*
  - *OL-3.1-R2.1 Program header information: Priority 1*
  - *OL-3.1-R2.2 Observation status information (and schedules themselves): Priority 1*
  - *OL-3.1-R2.3 Field information: Priority 1*
  - *OL-3.1-R2.4 Data from interferometer and single-dish organized by:*
    - *OL-3.1-R.2.4.1 position: Priority 1*
    - *OL-3.1-R2.4.2 subreflector state: Priority 1*
    - *OL-3.1-R2.4.3 polarization products: Priority 1*
    - *OL-3.1-R2.4.4 spectral channels: Priority 1*
    - *OL-3.1-R2.4.5 frequency bands: Priority 1*
    - *OL-3.1-R2.4.6 IFs: Priority 1*
    - *OL-3.1-R2.4.7 subarray including those of ACA systems (the TP Array and the 7m Array): Priority 1*

Requirements: Data Format (Cont.)

- *OL-3.1-R2.6 Total power (autocorrelation) data from antennas in single-dish mode : Priority 1*
- *OL-3.1-R2.7 Weights and/or data uncertainties: Priority 1*
- *OL-3.1-R2.8 States indicating special modes (such as ON/OFF positions), for:*
  - *OL-3.1-R2.8.1 OTF scanning: Priority 1*
  - *OL-3.1-R2.8.2 subreflector switching: Priority 1*
  - *OL-3.1-R2.8.3 frequency switching: Priority 1*
- *OL-3.1-R2.9 Flagging data or masks: Priority 1*
- *OL-3.1-R2.14 Processing history: Priority 1*

Requirements: Other

- *OL-3.2-R1 A variety of data formats must be supported by the Package:*
  - *OL-3.2-R1.1 The ALMA standard archival data format must be supported for input without loss of functionality or information (includes both interferometric and ACA single-dish data format): Priority 1*
- *OL-3.2-R1.2 Other standard formats as designated by the ALMA Project shall also be supported for both input and output without loss of functionality or information.*
  - *OL-3.2-R1.2.3 Single dish FITSs: Priority 2*
- *OL-3.2-R3 The Package must be able to handle, efficiently and gracefully, datasets larger than main memory of the host system: Priority 1*
- *OL-3.4-R1 Data produced by other interferometers and single dishes in similar observing modes shall be importable and processable if provided in ALMA standard archival data format or an ALMA supported data format: Priority 2*

Score Sheet of CASA Requirement

ID	Priority	Category	Description	Status
OL-3.1-R2.1	1	Data	Program header information	**Partly Available**
OL-3.1-R2.2	1	Data	Observation status	**Partly Available**
OL-3.1-R2.3	1	Data	Field information	**Available?**
OL-3.1-R2.4.1	1	Data	Position	**Available**
OL-3.1-R2.4.2	1	Data	Subreflector state	**Available, But...**
OL-3.1-R2.4.3	1	Data	Polarization products	**Available, But...**
OL-3.1-R2.4.4	1	Data	Spectral channels	**Available**
OL-3.1-R2.4.5	1	Data	Frequency bands	**Available?**
OL-3.1-R2.4.6	1	Data	IFs	**Available?**
OL-3.1-R2.4.7	1	Data	Subarray information	**Not Available**
OL-3.1-R2.6	1	Data	Total power data	**Available**
OL-3.1-R2.7	1	Data	Weights/uncertainties	**Not Available**
OL-3.1-R2.8.1	1	Data	OTF scanning state	**Available**

Score Sheet of CASA Requirement

ID	Priority	Category	Description	Status
OL-3.1-R2.8.2	1	Data	Subreflector switching state	**Available**
OL-3.1-R2.8.3	1	Data	Frequency switching state	**Available**
OL-3.1-R2.9	1	Data	Flagging information	**Available**
OL-3.1-R2.14	1	Data	Processing history	**Available, But...**
OL-3.2-R1.1	1	Data I/E	Import ALMA data without loss	**Not Available**
OL-3.2-R1.2.3	2	Data I/E	Import/export SDFITS without loss	**Partly Available**
OL-3.4-R1	2	Data I/E	Importable other telescope's data	**Partly Available**

Uncertainty in Scantable Definition

- IF with different number of channel
  - More specifically, spectral data and TP data
  - nChan is unique for Scantable, but we can store spectra that have different number of channel
- Polarization
  - POLTYPE and POLNO
  - Enumeration rule for POLNO (STPol)
    - e.g. 0: XX, 1: YY, … for POLTYPE=LINEAR
  - Is this correctly handled in reader/writer?

PKSMS2writer Failure Case

- IFNO is not 0 in the first row

- POLNO is not 0 in the first row

- Number of polarization is not identical

- Multi-beam
  - Several DIRECTION values with same TIME

- …

Scantable: Merits and Demerits

- Metris
  - Relatively simple
  - Dedicated for single-dish data
    - e.g. one-to-one relation between SPECTRA and DIRECTION
  - Smaller data size
- Demerits
  - No detailed documentation
  - Some uncertainty
  - Incomplete reader/writer
    - failure case
    - some attributes are hard coded

MS: Merits and Demerits

- Merits
  - Closer design to ASDM
  - Advantageous to combination with interferometry data
  - Multi-antenna acceptable (ALMA have four TP antennas)
- Demerits
  - Too complex
  - Not adequate for single-dish data
    - e.g. need a search to get DIRECTION for DATA
  - Larger data size

Data Size

- Test data: OrionS_rawACSmod (MS format)
  - Initial data size: 154MB
  - Includes a few optional tables for GBT
- Convert to Scantable (using sdsave task)
  - data size: 21MB
- Convert back to MS (using sdsave task)
  - data size: 132MB

In View of Users

- Coexistence of two data format in one software
  - Annoyance?
  - It may be OK if it is well documented
- Coexistence of two plotter in one software
  - Confusable?
  - Plotter should be merged in CASA if it is OK

In View of Developers

- Duplicate development
  - May arise the same issue in both interferometer side and single-dish side

- Plotter requirement
  - OL-2.2-R6
    - The look and feel of the GUI will be uniform throughout the entire package

Candidate Plans

- Scantable should be updated anyway…
- May be (at least) two choices:
  - Extend Scantable definition
  - Switch to MS
  - …

Plan 1: Extend Scantable Definition

- Design should be taking into account ALMA requirements
  - Information must be kept at least mandatory ones for single-dish observation/analysis
- Documentation
- Reader/writer must be updated

Plan 2: Switch to MS

- (Almost) all codes must be rewritten
  - Especially, getter and setter
- Interface change?
- A few features would be needed to recover undesirable redundancy in MS
  - e.g. DIRECTION information
- Need test on performance
- Is the MS definition enough?

Efforts

- Plan 1: extending Scantable
  - Less efforts than Plan 2
  - Total
    - Design (data format):
    - Coding:
    - Testing:
- Plan 2: switch to MS
  - More efforts than Plan 1
  - Total
    - Design (implementation):
    - Coding:
    - Testing:

Summary

- Overview of data format issue
  - Scantable is advantageous for single-dish data, but seems to be updated its definition to keep necessary information of ALMA data
  - MS has closer definition with ASDM, but is too complicated
- Candidate plans
  - Extend Scantable definition
  - Switch to MS