

Pipeline Requirements

F. Gueth, IRAM Grenoble

Real-time calibration pipeline (I)

007, 016, 017, 018, 024, 060

- Definition [3.6.1.1]:
 - atmospheric model
 - instrumental calibration: process pointing, focus, delay, etc.
 - calibration with no time-interpolation: process and store the results
 - calibration with time-interpolation: process, results to be used by Dynamic Scheduler, and QL Pipeline
 - target sources: just store the atmospheric calibration calibration

Real-time calibration pipeline (II)

007, 016, 017, 018, 024, 060

- Calibration modes are not yet chosen (cf. summary from Calibration meeting)
- Storing the atmospheric calibration in RT operations [6.3.2–R2]: is this required? an implementation issue?
- Distinction between two types of calibration (require or not a time interpolation) is less clear than advertized (020, 024, 060)
- What does “real-time” really mean? (016, 060)
- Terminology varies accross the document (018, 047, 048)

Real-time calibration pipeline (III)

012, 013, 016, 024, 064, 074, 075, 078, 093

Baselines, holographies, pointing models, beam properties measurements, etc, are not real-time tasks [6.3.4–R2 to R6]

→ not part of pipeline

→ but part of telescope calibration

The pipeline should be able to accommodate plug-ins for the handling of special observations like holography; absolute pointing model generations; baseline determination, etc. (074)

Need to detail these non-RT telescope calibrations, and their status: they do not belong to a specific project (093)

Quick-Look pipeline

007, 014, 017

- Definition [3.6.1.1]: monitoring of the observations; the QL pipeline is on the top of all other operations and does not perform any operation required by other parts of the system
- Distinction between Real-Time and Quick-Look operations is vague, merge them (014).

Science pipeline

007, 014, 018

- Definition [3.6.1.1]:
 - Science Calibration: process data acquired during the observing session that has just ended
 - Science Imaging: process all data observed for that programme
- Distinction between Science Calibration and Science Imaging is too fine; make sense in “waterfall” model, but not if self-calibration is used (014)

When should the pipelines be activated?

018, 021, 048, 049, 055, 081, 093

- Real-Time operations: after each observation [6.3–R1] \implies only after calib. obs. (049)
- Quick-Look: after RT or whenever requested [6.4–R1] \implies too often (081)
- Science Calibration: after reaching a break-point or at the end of observing session [6.5–R1] \implies too often (018), run it more often (021)
- Science Imaging: after the Science Calibration; results are preliminary after observing session, final after completion of programme [6.6–R1] \implies need to update archive? or to have preliminary + final archives? sync. with archive copies (RSC)?

Pipeline scripts

026, 028, 035, 036, 040, 106

[6.2–R3] The Pipeline shall operate through readable and comprehensible data reduction scripts. Those scripts shall be automatically generated from templates, on the basis of the observing mode being use.

- Is the second sentence an implementation issue? (026, 028, 036, 040)
- Script in which language? (028)
- Pipeline scripts: provided by ALMA project (i.e. standard)? by the PI? both? (035, 106)

Reversability of processing

026, 029, 031, 037, 044

[6.2–R6] Sufficient recording of operations performed and control parameters shall be carried out so that any step can be reversed and redone if needed, without recourse to repeating an entire series of previous operations or resorting to a copy of the dataset at the intermediate state.

- *This requirement is unreasonable (...)* (026)
- *...is very general requirements.* (029)
- *I think the general aim of this requirement is necessary (...) However, (...) I consider its strict application too complex processing.* (031)
- *Redoing is not the same as reversing. (...) I think this requirement is close to requiring the ability to undo operations, which is very expensive.* (037)
- *Yes; but add : ... reversed and redone, taking into account the order of operations and the fact that some of them are non-commutative.* (044)

Interaction with real-time system

012, 016, 059, 062, 063, 065, 072, 076

- Results from Telescope Calibration are “passed” to the system, or “made available” (065), or “archived” (016, 059, 072)? [6.3.2–R3, 6.3.4–R1, 6.3.4–R2, 6.3.4–R6]
- *In general, it would be better if the authors indicated how they want quantities used, rather than to which (as yet undefined) “part” of the system they should be sent. (012, 062, 063, 076) [6.3.2–R1, 6.3.2–R3, 6.3.4–R1 to R6]*

Requirements too prescriptive?

007, 038, 054, 087, 089, 114

- *There is a tendency to be overly prescriptive in describing how the reduction is to be done. I think that our best practices will evolve as ALMA comes online and the requirements should reflect this. (007) [6.2–R9, 6.3.2–R3, 6.4.2–R3, 6.4.2–R4, 6.6.1–R9]*
- *What is described is the current best practice for mm arrays. This may not be the best practice for ALMA. Does one want to bind ALMA to work this way? (054) [6.3.2–R3]*

Is this only a matter of wordings, or should some requirements be replaced by a reference to “best practice in the ALMA epoch” ?

Requirements too ambitious?

034, 103, 110, 114

- *The goal of being able to process all data from the array in standard modes is too ambitious (034, 007) [6.2–R1]*
- *I suggest that having "several [deconvolution] algorithms running in parallel" is better done as an offline, rather than an online task. (103) [6.6.1–R7]*
- *(...) there is a combinatorial explosion possible (weighting x deconvolution x other parameters) (110) [6.6.1–R6 and R7]*
- *This is really a call for more development of automated methods for identifying and removing continuum. In the absence of any known method, one cannot require that it be used! (114) [6.6.1–R9]*

Is this a matter of setting priorities, or are some requirements not realistic at all?
Provision for future algorithms/methods developments?

Internet?

80, 82, 100, 105

[6.4–R2] results of the QL Pipeline must be available via the Internet

[6.6–R4] results of the Science Pipeline must be available via the Internet

- *How much data should be made available to the PI's over the Internet? When? In near real-time? Suppose the PI's aren't available or are asleep? (080)*
- *Whether the Internet is an appropriate delivery medium for Science Imaging results regardless of their size seems to me debatable. If a PI has been waiting for six months for a project to complete, why can't he/she wait a few more days for the delivery of a DVD? (100)*

Miscellaneous

- Is the Pipeline built using the Off-line package? (009)
- Single-Dish vs. Interferometer operations: what should be mentioned? (015)
- The case of multiple sources and/or multiple frequency setups in the same project is not mentioned (021, 023)
- What does “data-driven” mean? (026) [6.2–R2]
- Separation pipeline–ALMA system (030)
- Need more details concerning mosaic and self-calibration modes [6.4.2–R4, 6.6.1–R9] (085, 088, 090)
- Can the Pipeline make use of observations which are not part of the project? (095, 108)
- If the flux scales are different, what is to be done? [6.6.1–R3] (102) i.e. what is the Pipeline answer if there is a problem during data reduction that precludes producing a reliable result?