

Comments on: Software Science Requirements and Use Cases

July 11, 2002

1 Comments on Pipeline Section (3.6) and replies

prepared by Frederic Gueth and Peter Schilke

Comment 003 by MasatoshiOhishi

General Comments:

In the ITU (International Telecommunication Union), we discuss possible frequency allocations for Satellite Operators in the 40 GHz range and around 94 GHz. I think we need to consider how to monitor and mitigate interferences due to space-to-earth direction transmissions from satellites. It is necessary to make real-time monitoring to avoid unwanted emissions from satellites.

Reply:

An addition requirement (6.4.1.-R3) has been added in that sense.
"The Pipeline shall be able to detect interference,
e.g. due to communication satellites operating at or near the observed frequencies, flag the data accordingly, and trigger alarms if necessary. (Priority 3)"

Comment 005 by RayPlante

Hi Robert,

Unfortunately, my first comment is about a requirement that is (I think) missing, so I feel compelled to break from the requested format in Part 1. Part 2 refers to existing requirements. I hope my discussion is not too lengthy to be helpful. Your committee's response need not be as detailed.

1. User Access to Software

I feel an explicit requirement is necessary about the usefulness of data in the archive to end users (astronomers); something like,

"Astronomers shall have [free?] access to software necessary for processing all data products available from the archive."

This may seem obvious (hopefully, I didn't overlook such a statement). It is certainly implied by other requirements in sections 1, 3, and 4; however, I think it's important to spell this out as I think it affects how these other requirements are met (and the cost of doing so). Some of my other comments below hinge on an implicit assumption of this nature.

You probably should go even further. Here are some possible riders:

- * "In particular, the astronomers should have access to all software necessary for duplicating or redoing the [reversible] processing carried out by the pipeline"
- * "...on their local computing platforms"
- * the data reduction scripts are locally runnable

To the extent that these are true, they should be expressed explicitly. The extent to which they need not be true, that should be stated as well--for example,

- * restriction: limited to data products in the Observational Archive.
- * possible exceptions: Atmospheric Model, Quick Look operations
- * products made available by astronomers (e.g. final images) resulting from local processing are exempt from this requirement.
- * some projects may require high performance computing to perform the processing; performance requirements (execution time, memory, disk space, etc.) should not be considered when meeting this requirement for such projects.

Reply:

You should look into the requirements for off-line data processing. Look in particular OL-1.1-R5, OL-1.2-R8. The pipeline will also be runnable at places where the archives are kept; these are known as "Regional Support Centers", and are foreseen by the ALMA project.

Comment 007 by TimCornwell

Apologies for being late. I hope that you can consider my suggestions. I appreciate being asked to review these requirements. I found it helpful for our own work here on end-to-end processing of data from NRAO telescopes. I have a few global concerns:

- The distinction between the various pipelines is insufficiently motivated.

Reply:

We have tried to clarify the distinctions. The Science Calibration and Science Imaging pipeline were merged into a single "Science Pipeline". The real-time calibration and the quick-look pipeline have obvious inter-relations, and could therefore be merged into a single "real-time pipeline". However, we kept the distinction, in order to emphasize which operations are necessary for the system running, and which are only performed for monitoring. The distinction is more an architecture issue than a requirement issue. In fact the Analysis group has taken the view of integrating the (real-time) Calibration Pipeline in the Instrument Operation Subsystem, not the Data Reduction subsystem.

- The goal of processing all standard modes should be qualified by the attachment of quality measures.

Reply:

Accepted. An additional requirement (6.5-R5) has been added for that purpose.

- 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.

Reply:

We have added a number of "provisions" to make it clear that only the current best-practices are described in this document. On another hand, the ALMA pipeline should be developed during the coming years, and the SSR document should be as prescriptive as possible, in order to be useful for the developers.

Comment 009 by WimBrouw

Comments on Pipeline part of ALMA SSR and Use cases -- version 4

Introduction

Add in the introduction that the Science calibration and Imaging will be a sub-set of the off-line data reduction package on the one hand; and that the real-time calibration will be available in the off-line data processing package.

Reply:

This would imply that the Pipeline be built with the Off-Line package. While this may certainly be true, it is felt that this is not a requirement for ALMA software. The requirement is that all operations done by the Pipeline can also be done off-line.

Add that pipeline processing will have parallel streams, with a mixture of synchronous and asynchronous operation

Reply:

This is an implementation issue.

Question (in principle for Use cases): will the measured WVR be archived? If not, calibration based on WVR should not be included in off-line package.

Reply:

Yes, the WVR data will be archived. See Req. 7.2-R2 (new version).

Comment 012 by JoeSchwarz

p. 26, 3.6.1 "NOte" --> "Note"

"Occurrence" is misspelled as "occurence" everywhere in this section

3.6.1.1 "make results available to the dynamic scheduler" -- this assumes a certain division of the system into items that include a "dynamic scheduler." Since things like preliminary phase rms will be used to adjust target and calibrator dwell and cycle times, the requirement as expressed would not address this use. 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.

Reply:

Accepted. We have replaced "Dynamic Scheduler" by "dynamic scheduling system". This part of the system is described in another part of the SSR document; the list on required inputs to the dynamic scheduling algorithm is given in Req. 4.0-R3. Some of those input parameters are supplied by the Pipeline, which we have indicated in the Pipeline requirements. This should be cross-checked with the dynamic scheduling requirements.

Are baseline and holography calibrations really "(quasi) real time"? Robert Lucas indicated to the software group that baseline calibrations might be deferrable to a time after scientific observations corresponding to the (new) antenna configuration had been completed--in some cases.

Reply:

Accepted. Baseline and holographies are not quasi-real time operations -- except possibly in some cases, as high-frequency work, where the dynamic scheduling may have to take into account the baseline quality. As suggested by another reviewer, we could consider such operations as "plug-ins". The Requirements 6.3.4-R2 to R5 have been rephrased accordingly.

Again, "make results available to the sequencer" doesn't say what these results are to be used for. "Sequencer" is not a meaningful concept at the requirements stage.

Reply:

Accepted. "Sequencer" has been replaced by the very general term "real-time control system". Some parameters derived by the Pipeline (e.g. pointing offsets) shall actually be used by the control system in order to update the instrument parameters.

In the same vein, requiring Quick Look Operations to be done during and after an "observing session" doesn't give developers the information they need to make these operations useful. Who is to look at the results? When? What kinds of corrective actions should be made possible by these results?

Reply:

The QL Pipeline is there to allow the ALMA staff to monitor all observational/astronomical informations that can provide a valuable information to detect any misbehaviour of the instrument. The results are also provided to the observer (PI) by means of a web page (see 4.0-R9 and UC 4.5.2).

Comment 013 by MasatoshiOhishi

p.26 3.6.1.1 Real-time Calibration Operations

In Telescope / Array Calibrations, although I could be wrong, do we really need to reduce the holography data ? See also subsection 3.6.3.

Reply:

See comment 12.

Comment 014 by TimCornwell

Page 26, 3.6.1.1 Pipeline operations

The distinction between Science Calibration and Science Imaging is too fine to be worth building in at this level. While calibration and imaging used to be considered as separate items in a waterfall model, a more modern view is that the two are intertwined via the concept of self-calibration.

Reply:

Agreed. The two sections are now merged.

I am similarly dubious about the distinction between real-time calibration and quick look operations. I'd recommend the distinction be between Real-time and Post-observing Processing.

Reply:

Cf comment 007. The difference here is in the motivation (real-time calibration is for automatic detection of problems, quick look is for humans), the data being processed (well known calibrators vs science objects) ...

Comment 015 by WimBrouw

p26 3.6.1.1: Most of the following details in this section are valid for all types of operations. Iso saying 'relates to interferometric', why not have an indicator to exclude the few items that are not valid for single dish or light-bucket operations.

Reply:

Agreed.

Comment 016 by WimBrouw

p26 3.6.1.1 RT cal operations
 - Some of these are 'real' RT, i.e. they have to be done before observing can proceed, since errors will remove information (like focus; delay; offset pointing); others are not (holographics; baseline tests)

Reply:

Agreed.

- The Atmospheric model is RT; some of the astronomical ones could be (atmospheric calibration idf that is the one based on the atmospheric model); others are definitely not (bandpass calibration; prelimin phase and ampl cal)
 - Is the 'preliminary phase rms' purely based on atmospheric information? If not, what is its purpose for the dynamic scheduler (since the phase noise will depend on baseline; ampl; atmosphere, ..

Reply:

The phase rms are computed from the calibrator measurements. They are used to estimate the seeing, and are an input for the dynamic scheduler.

- Can it be indicated which one are 'await result RT', and which are not?

Reply:

The list given in 3.6.1.1 is mainly to give a overall view of each group of operations; the precise definitions are given in the requirement list.

- Indicate that this processing should include, but not be limited to the list given

Reply:

Agreed.

- Should results be made available to sequencer (and/or scheduler) or should they be archived, and be available through archive?

Reply:

This is an implementation issue.

Comment 017 by WimBrouw

p26

3.6.1.1 QL operations

- Occurrence: after observing session, and during observing session at tbd intervals
- Monitoring-- done through archive access of cal data?
- Quick calib -- apply should be before resample/integration; also what is the calibration done here -- oh, I see apply only; not do. Maybe some terminology to indicate active and passive use of the verb 'to calibrate' should be used. Maybe use here 'correction' rather than 'calibration'
- Quick imaging: rather than 'no deconvolution'. I would think 'subset of data' is more important (specified beforehand e.g. for spectral line to integrate channels with expected result
- Display tools: display 'information about' current observations. Display of monitoring information and non-imaged data (e.g. baseline-time amplitude images) and statistical extractions are more important than 'images' (especially for non-SNR=100 single sources)

Reply:

The list given in 3.6.1.1 is only to give a overall view of each group of operations; the actual definitions are given in the requirement list. In order to avoid ambiguities, the list in 3.6.1.1 has been simplified.

Comment 018 by JoeSchwarz

p. 27, Science Imaging Operations

Accessing the archive for previous observations, producing and archiving deconvolved images are expensive to do "after completion of observing session." My understanding from the SSR discussions was that these would be done,

at most (and this is what is implied by requirement 3.1-R13)
 after breakpoints and (of course) when
 the project (or at least observations of the target) is
 complete. This requirement is also in conflict with Use Case 4.5.3 on p. 73.

Reply:

Accepted:

After discussion in the Granada Meeting we have decided to separate occasions when full science data processing is performed, from the break points and the end of sessions. The PI will specify in the Observing Tool when Monitor Points are requested: after each scheduling block, after each end of session (provided a minimum observing time is reached), or at break points only. The necessary computing resources will be evaluated by the Observing Tool and this evaluation will be available to the reviewers.

3.6.1.3 It seems as though "instrumental" calibrations were called "telescope" calibrations in 3.6.1.1. The change in terminology is confusing.

Reply:

Agreed.

Although bandpass calibrations do not require a time interpolation, and can therefore be "immediately derived and stored, to be applied to all following observations," 3.6.1.1 says it's "derived or applied" after completion of observing sessions. This seems inconsistent to me, and is another example of how the requirements can get confused when they contain implementation considerations instead of real requirements.

Reply:

Agreed.

Comment 019 by TadafumiTakata

p27 3.6.1.1 Pipeline Operation
 Science Imaging Operation

Comment

Accurate positional calibration may be needed in producing deconvolved images, which are most popular data for general archive users. They have to include coordinate information like FITS WCS headers which can be interpreted by image browsers to be provided to ALMA users etc.

(Please ignore this comment if it is already included in the pointing and/or instrument calibration before producing science images.)

Reply:

Agreed.

Comment 020 by TimCornwell

Page 27, 3.6.1.3 Calibration

I don't think that one can draw a line between items requiring interpolation and those not. All items must benefit from interpolation (or modeling in time and space). Another important and related factor is quality control: one needs access to a time-series to allow discovery of errant values. The key point here is whether one must wait for all relevant values before making a prediction of calibration values at some point in time and space.

Reply:

The Science Pipeline should use the best technique available, including checks of any time-variation of all calibration parameters. However, this would make no sense for the Quick-Look pipeline, which should use the simplest method.

Comment 021 by WimBrouw

p27 3.6.1.1 Science cal operations (science correction(?))

- Why after completion of session? Why not also require option during session. I could think of a program; and hence an observing session, observing many individual fields as snapshots for getting to know some parameters of a series of 'sources'. Run it in between
- derive and/or apply: bandpass (not a curve according to other places but fixed; a session could have many bands and places in bands; and hence many bandpasses
- what is 'final' phase and ampl calibration: cannot be precise enough in pipeline with 'limited or no deconvolution and selfcal). Or is 'final' as far as the pipeline is concerned?
- flux scale is that derived here?

Reply:

The list given in 3.6.1.1 is only to give a overall view of each group of operations; the actual definitions are given in the requirement list. In order to avoid ambiguities, the list in 3.6.1.1 has been simplified.

Comment 022 by WimBrouw

p 27

3.6.1.1 Science imaging

- should 'produce temperature-calibrated visibilities' be in previous step? (btw why day 'uv tables', is implementation
- is 'deconvolved' true here? If so; there should be an extensive (self-)calibration step in science calib operations to be able to derive that.

Reply:

See comment 21.

Comment 023 by WimBrouw

p27

3.6.1.2

- astronomical 'source': a project can have many sources (list) or a large [mosaiced] field. Mention that source here is not the normal single, limited source.

Reply:

Rephrased.

Comment 024 by WimBrouw

p27

3.6.1.3

- instrumental: indicate which ones (as mentioned above) are part of the RT-wait set; and which ones are not

- no time interpolation: I am not sure it is a good idea to give these examples as 'time-invariant' calibrations. There are quite a few schemes where it is easily possible to have to interpolate e.g. bandpass (filter poles can be very T sensitive; if receivers are going to apply diurnal Doppler); bandpass could also have a phase error. I would leave it open. Also: bandpass and pol are non time-critical:

Also, indicate listsinclude but are not limited to)

Reply:

See comment 20.

Comment 025 by WimBrouw

p28. Finally ... average of observed... I.e. leave out 'the', and indicate that the average will be done outside the pipeline operations (there could be contaminated channels e.g.)

Reply:

The pipeline shall produce a pseudo-continuum measurement, using astronomers' inputs. The final average is to be done by the astronomers, off-line.

Comment 026 by JoeSchwarz

p. 29, 3.6.2-R2 What does "data-driven" mean here?

Reply:

It means that the Pipeline should run automatically as soon as new data is acquired, and that operations to be performed should be determined by the nature and purpose of the data, as indicated by header information associated with it.

3.6.2-R3 The meaning of, and need for, "templates" isn't clear. Isn't this more of a design issue for the people doing the observing preparation tools?

Reply:

Agreed. The second part of the requirement has been removed.

6.2-R4 I understand a "pipeline" as something that, once started, runs automatically. In contrast, "tools" are usually instruments for doing some task that one wants to direct interactively. "Automatic flagging tools" sounds like a contradiction to me. I think you mean that certain data should be flagged automatically.

Reply:

Rephrased.

6.2-R6 This requirement is unreasonable, in my opinion. You can't repeat a series of previous operations and you can't resort to a copy of the dataset at an intermediate state. "Sufficient recording..." then means that results of every step have to be saved -- but not as a "copy of the dataset". Maybe I misunderstand, but I think that a straightforward interpretation of what's written here will produce a very heavy and expensive system. I think the kinds of "steps" that can be reversed and redone should be spelled out in more detail.

Reply:

This requirement has been removed.

6.2-R9 I don't understand. Doesn't the quantity to be calibrated determine whether it's baseline- or antenna-based? (Maybe this is a stupid question, but the Aperture Synthesis Summer School isn't until after the deadline for comments.)

Reply:

In some cases, antenna-based calibration may fail, in which case baseline-based calibration is to be used.

Comment 027 by MasatoshiOhishi

p. 29 3.6.2 Pipeline general requirements

In 6.2-R4, the degree of interference should be added as a condition to flag data.

Reply:

Agreed.

Comment 028 by PrebenGrosbol

p.29 6.2-R3

'through readable and comprehensible data reduction scripts'
Sounds good but I think there are additional points, namely: a) if observing scripts exists it would be better to use the same control language for both observing and reduction scripts, and b) in order not to be too dependent on a specific pipeline engine the script language must be independent of any specific processing system (ref. Reduction Blocks for the VLT). That is the script should provide the flow control and specify the reduction tasks to be executed while the engine just should execute these tasks as best it can. Further, there is a verification issues (like for observing scripts) if users are allowed to change them.

Reply:

The second part of the requirement has been removed, in order to be less prescriptive.

Comment 029 by PrebenGrosbol

p.29 6.2-R6

'Sufficient recording ... shall be carried out so that any step can be reversed' is a very general requirements. Some processing is difficult to reverse and often it is simpler to save some intermediate results. Recording all reduction steps is clearly needed but already covered by 6.2-R5.

Reply:

The requirement has been removed.

Comment 030 by PrebenGrosbol

p.29 3.6.2 Pipeline General Requirements

I am missing some requirement on that the engine used for pipeline processing should be well separated from the ALMA system i.e. it should be possible to replace the pipeline engine without any significant change in the ALMA software. Although this partly is a design concern since one would not like to have the ALMA software depend on an alien, uncontrolled package, it is also a science concern. One would like to enable anybody to contribute useful pipeline modules to ALMA and not limiting them to software written in a specific system. Further, with a lifetime of decades for ALMA it is likely that it will outlast one or several pipeline engines.

Reply:

We believe this is a fair design concern, based on experience.

We propose to add the general requirement:

The choice of a specific data processing package to provide the main pipeline functionality should not impose unnecessary constraints on the rest of the ALMA software system, in particular on the manner in which the data is stored and handled, as in the future new or better functionalities may be provided by newly developed data processing systems.

Comment 031 by RayPlante

p. 29, 3.6.2, 6.2-R6: (Reversability of processing)

I think the general aim of this requirement is necessary: we want the ability to remove or substitute any part of the processing. In particular, we want to be able redo it with small modifications to the parameters. However, the wording of this requirement concerns me when I consider its strict application to complex processing; some clarification may be helpful.

This requirement places a corresponding requirement on the off-line software (assuming point 1 above), the data formats used, and on what products must be stored in the archive. Should one be able to back up an arbitrary number of steps in processing ("with out resorting to a copy at an intermediate state")? What constitutes a step? Suppose target source has been self-calibrated with multiple loops; does reversability mean that we have to keep the generated gain table and image models for each loop? Taken to its extreme, this requirement could be pretty costly:

- * if all datasets required for step-wise reversability, then the archive must organize and label these products in a way that they make sense to the user.
- * the data formats must retain sufficient information for backing up.
- * for every processing step implemented, its reversal must also be implemented. The cost is greater if this capability must be available to the end astronomer (see 1 above).

If we only mean to apply this requirement to the extent already supported by current packages, we may be okay here. If we only

require needing to make one step backward at any given step in the chain, we might be okay.

The cost is of particular concern if it's an unnecessary one. Thus, I'm curious as to the intent of the clause "without...resorting to a copy of the dataset at the intermediate step." Redoing processing by going back to an intermediate product is often the easiest way to "back out" (as opposed to reversing multiple steps in turn). If the processing script is available (which is required) and the assumption from 1. (above) is true, then this is straight forward.

This requirement could be clarified in the context of the definition of ranked data products in the archive. For example:

```
Level 0 -- raw data: has no/certain/all real-time calibration
operations applied
Level 1 -- calibrated data: has science calibration applied
Level 2 -- deconvolve images: has Science imaging operations
applied
Level 3 -- final, user-supplied (locally processed) data products
```

Then, say, if one starts with the data products at a given level and apply the scripts that produce the next level stepwise, it should be then possible step backwards at any point in the chain.

By the way, defining ranked data products will help the user understand what they need from the archive. They can easily choose what level of processing they want to accept and know what products they, therefore, need to retrieve.

Reply:

The requirement 6.2-R6 has been removed.

Comment 032 by RayPlante

p. 29, 3.6.2, 6.2-R8:
 "Sequencer" -- you probably need an entry in the Glossary for this. I was a little unclear about what it does. According to my pdf reader, this is the first use of this term in the document; although, its function may be made clear in the Use Cases (which I did not study as well).

Reply:

"Sequencer" has been replaced by the very general term "real-time control system".

Comment 033 by TadafumiTakata

p29 6.2-R1.1
 Comment

Is there any need to have a feed-back (like result images and so on) from this operation ? If so, some tools for helping the process to reflect the result to ALMA archive should be necessary.

Reply:

The concern is that if one allows PIs to contribute the results of those special projects into the Archive, then ALMA has no control whatsoever over the quality of the data. It is understood that, to be usable for archival research (VO), the quality of the data has to be evaluated in a well defined way.

Comment 034 by TimCornwell

Page 29, Pipeline General Requirements

I think that the goal of being able to process all data from the array in standard modes is too ambitious, certainly so for a telescope yet to be built. For the EVLA, VLBA, and GBT pipelines, we are planning to attach a quality measure to pipeline results. This adds an important level of qualification. A division might be:

Meets all quantified scientific goals

Meets some quantified scientific goals

Meets none of the quantified scientific goals

A priority 2 requirement could be added that the pipeline must process all data from all standard modes to the highest level of quality. The priority 1 requirement is to process all standard modes and attach a quality measure.

Reply:

Excellent suggestion. A requirement (6.5-R5) has been added in that sense.

The requirement that the pipeline not constitute a bottleneck is a fine sentiment but I'm not sure who it is directed to? - TAC, operations?

Reply:

Computing engineers, I guess.

Comment 035 by TimCornwell

Page 29, 6.2-R2:

Some default parameters (cellsize, field of view, calibration methods, etc) should come from a standards database that is under change control.

Reply:

We believe that those parameters should be provided at proposal preparation phase, or defaulted at this occasion (so that the user may check their values if needed).

Comment 036 by TimCornwell

Page 29, 6.2-R3:

The second sentence is an unnecessary implementation detail. For NRAO pipelines, we are generating scripts from production rules encapsulated as make files. This is arguably superior to templates. In any event, it's not necessary to state how the scripts will be generated.

Reply:

Agreed. The second sentence has been removed.

Comment 037 by TimCornwell

Page 29, 6.2-R6:

Redoing is not the same as reversing. I'd recommend removing the reversing part. To redo, one simply needs checkpoints. To reverse, one needs much more. I think this requirement is close to requiring the ability to undo operations, which is very expensive.

Reply:

The requirement 6.2-R6 has been removed.

Comment 038 by TimCornwell

Page 29, 6.2-R9:

I think this requirement is not necessary. It is true that antenna-based calibration is better than baseline calibration for effects that really are antenna-based. However, it almost certainly is true that physical modeling of antenna phases by e.g. time and space parameterized phase screen is superior to antenna-based calibration. Perhaps the requirement should state that "Best practices" must be following in calibration?

Reply:

Requirement rephrased.

Comment 039 by WimBrouw

p29 6.2-R1.1 This holds for science data; I would suggest that the pipeline always handles 'active calibrator' data

Reply:

The calibration pipeline should handle these data in the standard way (pointing, focusing, checking phase stability) in order to ensure that the requested observing conditions are met.

Comment 040 by WimBrouw

p29

R3: shall operate through 'automaticly generated ' readable ... Drop the second part (implementation)

Reply:

Agreed.

Comment 041 by WimBrouw

p29

R4: 'step' is undefined; say: the pipeline shall include automatic flagging of data. Do not use 'discard' (bypass? not use).

Reply:

Requirement rephrased.

Comment 042 by WimBrouw

p29

add: R4.1: Flagging should be multi-level to enable selctive re-use of automatically flagged data; or the automatic flagging must be reversable in the off-line stage using identical algorithms

Reply:

It is already mentioned in this requirement that the flagging must be reversible at the off-line stage.

Comment 043 by WimBrouw

p29

R5: is 'output' archiving (should be)?

Reply:

Yes.

Comment 044 by WimBrouw

p29

R6: yes; but add : '.. reversed and redone, taking into account the order of operations and the fact that some of them are non-commutative, and have to be 'done and redone' ...'

Reply:

6.2-R6 has been removed, because it was felt it was not a requirement for the pipeline operations.

Comment 045 by WimBrouw

p29

R7 and R2 are mutually exclusive. Maybe R7 could be rephrased slightly?

Reply:

Rephrased.

Comment 046 by WimBrouw

p29

6.2-R9: talking about Amplitude and phase only?

R9: could you give one or more examples where baseline calibration (of what) is required?

Reply:

In some cases, antenna-based calibration may fail, in which case baseline-based calibration should be used. Baseline-based calibration is required only for effects that are baseline-based: e.g. baseline based amplitude calibration is needed when atmospheric decorrelation during the basic integration time is important.

Comment 047 by JoeSchwarz

p. 30, 3.6.3.1 "Astronomical Calibration: Atmospheric Model" -- this terminology is inconsistent with that of 3.6.1.1, where "Atmospheric Model" and "Astronomical Calibration" are listed as separate categories.

Reply:

Title of 3.6.3.1 changed.

Comment 048 by MasatoshiOhishi

p. 30 3.6.3 Real-time Calibration Operations

In 6.3-R1 it says, "The real-time Calibration Operations shall be activated AFTER each observations." However in page 26, in subsection 3.6.1.1, it says, "Real-time Calibration Operations occur (quasi) real time." I feel these are inconsistent.

Reply:

They are not inconsistent, because the term "observation" is used according to its definition given in the SSR document, p. 10: it actually represents a short integration period.

Comment 049 by PrebenGrosbol

p.30 6.3-R1

'... after each observation' It sounds too strong for me. I would have expected that only after calibration observations the calibration Operations are done. Normally, calibrations would be done after each science observation which then would trigger the calibration operation but the current requirement is stronger.

Reply:

If the last observation is a science one, then the calibration pipeline has nothing to do. If it is a calibration observation, then several operations have to be performed. The requirement has been modified for clarity.

Comment 050 by PrebenGrosbol

p.30 6.3.2-R2

'convert the raw data into temperatures, or, alternatively, store the conversion factor' There seems no reason to give an alternative - either one or the other. I would prefer the latter since this would store the raw, acquired data with the factor and not apply it.

Reply:

Requirement removed.

Comment 051 by RayPlante

p. 30, 3.6.3, 6.3-R2:

"Whenever the results ... allows to identify" ==>
 "Whenever the results ... allow one to identify"

Reply:

OK

Comment 052 by TadafumiTakata

p30 6.3.1-R3

Comment

The resultant parameters of model calculation should be referred by users (including astronomers) via "The Data Extractor Tool" or so for evaluating how the data are affected by this parameters.

(I think it is already included in enviromental information extraction.)

Reply:

The results of the atmospheric model are to be archived, and should thus be accessible later on.

Comment 053 by TimCornwell

Page 30, 6.3-R2:

I think this should be priority 2. It's going to be very hard to do from the real-time calibration. An equivalent goal for the post-observing processing is possible and should be priority 1.

Reply:

Agreed.

Comment 054 by TimCornwell

Page 30, 6.3.2-R3:

This seems too prescriptive. 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?

Reply:

Requirement rephrased.

Comment 055 by WimBrouw

p30

6.3-R2 '...after?? ..' RT calib is necessary before some data can be taken (atmospheric model). Certainly an observing session should also end with one.

If you do flag; multi-level flagging is even more important.

Reply:

For flagging: see requirement 2.3-R2

Comment 056 by WimBrouw

p30

6.3.1-R1 ...The prediction will be based on measured atmospheric data, including, but not limited to:

Reply:

Rephrased.

Comment 057 by WimBrouw

p30

R1: what is 'line-of-sight'? on a per antenna basis (what I would assume, certainly for higher frequencies and longer available baselines); and does it assume some isotropy and/or weighting over the HPBW FOV?

Reply:

Agreed: this is done for each antenna independently (though some consistency between neighbouring antenna could be checked). There is no planned method to measure variations of atmospheric properties inside the antenna primary beam (may be infrared sensors could be used?). However most of the absorption occurs in the low atmospheric layers where all line of sights inside the primary beam see the atmosphere averaged over the same near-field antenna pattern.

Comment 058 by WimBrouw

p30

6.3.1-R3 already covered in R1? If not, what is different?

Reply:

R1 is a general requirement that a powerful atmospheric modelling be available. R2 and R3 are indicating two crucial uses of the atmospheric model.

Comment 059 by WimBrouw

p30

6.3.2-R1: 'store the results'? or 'archive the results'

Reply:

The verb 'archived' may suggest that the Pipeline writes the results listed in this requirement directly in the Archive. This is actually an implementation issue. What matters here is that those quantities are available for further Pipeline operations -- they could thus be written in the Archive at a latter stage, together with several other values. We therefore prefer the term 'stored', which is more general.

Comment 060 by WimBrouw

p30

R1.*: a time-constant indication is necessary here. It is a large variety of items (and again the list is probably not complete over life-time of instrument). I think 1 and 3 are per observation normally (or even longer). R2 could be fast varying?? (probably depend on length of observation as well); R4: is that in bore-direction (is really long-term to determine correctly). The polarised antenna pattern is something measured annually maybe.

Reply:

The software should be able to reduce all those observations. The frequency at which such calibrations are performed is an operation issue, which will largely depend on the observing modes, and on the expertise acquired using the ALMA. Hence, we prefer not to give such timescales in the software requirements.

Comment 061 by WimBrouw

p30

6.3.2-R2: alternative is incorrect. Tsys should always be stored (=archived). The alternative is the deferment of the conversion, not the storage.

Reply:

Agreed. This requirement has been removed.

Comment 062 by JoeSchwarz

p. 31, 6.3.2-R3.4 "...passed to the Dynamic Scheduler..." Again, it is difficult to understand how this data is to be used. Scheduling decisions are typically made on the timescale of an SB (nominally 15 minutes-1/2 hour, in order to take account also of the time needed to bring a new receiver band to a ready state). So what's needed here? Averages? Instantaneous values? Predictions of how these values will evolve over the next 1/2 hour?

Reply:

Cf. comment 12.

Comment 063 by JoeSchwarz

p. 31, 6.3.4-R2, How are the baseline calibrations to be used by the "Sequencer"?

Reply:

The "sequencer" (= the real-time control system, cf comment 12) shall use the antenna locations to compute the delays to be applied on each baseline by the correlator.

Comment 064 by MasatoshiOhishi

p. 31 3.6.3.4 Telescope / Array Calibration

In 6.3.4-R4, do we really need to reduce the holography data real-time ?

Reply:

No, not real time. Cf comment 12.

Comment 065 by PrebenGrosbol

p.31 6.3.3-R1

'... and pass the results to' For me 'pass' suggests something active i.e. the pipeline explicitly sends the information. I would prefer 'made available' only.

Reply:

Agreed.

Comment 066 by PrebenGrosbol

p.31 6.3.3-R2

Same comment as for 6.3.2-R2

Reply:

Same answer.

Comment 067 by PrebenGrosbol

p.31 6.3.4-R1

'must be passed or ...' Same comment as for 6.3.3-R1

Reply:

Agreed.

Comment 068 by TimCornwell

Page 31, 6.3.4-R2:

"baselines" is jargon. I'd change this to "antenna locations".

Reply:

Agreed.

Comment 069 by WimBrouw

p31

6.3.2-R3.3 : How can you have different parameters for the correction that is done on a fast basis before integration over samples?

Reply:

The idea is to choose the best conversion coefficient when the observation has been repeated using a few values. This may not be necessary.

Comment 070 by WimBrouw

p31

R3.4: what do you gain by doing this operation per baseline??

Reply:

It is useful to do amplitude calibration per baseline in case of amplitude decorrelation.

Comment 071 by WimBrouw

p31

6.3.3: Drop; or indicate again the earlier comment

Reply:

6.3.3-R2 is removed.

Comment 072 by WimBrouw

p31

6.3.4-R1: why explicitly made available to sequencer (earlier comment). Let sequencer determine what it needs: just archive/log it.

Reply:

Cf comment 12 on dynamic scheduling.

Comment 073 by WimBrouw

p31

6.3.4-R1.1: indicate (or in a more general comment) if it handles on offset pointing here (as I would assume).

Reply:

Rephrased.

Comment 074 by WimBrouw

p31

6.3.4-R4: should pipeline handle this? Why?

Maybe make it more generic:

' The pipeline should be able to accomodate plug-ins for the handling of special observations like holography; absolute pointing model generations; baseline determination etc'

Reply:

Excellent suggestion. Requirement rephrased.

Comment 075 by JoeSchwarz

p.32, 6.3.4-R5 Is the derivation of the primary beam properties from planets, etc., really a "real-time" operation?

Reply:

No, just like holography or absolute pointing models, these are telescope calibration operations to be performed regularly, but not to be reduced in "real-time". The requirements have been rephrased for clarity.

Comment 076 by JoeSchwarz

p. 32, 6.3.4-R6 "...passed or made available to the Sequencer..." Once again, we need to know how these results are to be used, and on what timescale. Even if we accept that there will be a "Sequencer" in the system, we have no "Sequencer Requirements" chapter (and I hope we don't write one!), so we need to know *what* is to be done with pointing, focus, skydip, etc. Are they to be used to correct the control system's pointing model so that the antenna points where it's supposed to, or rather to correct the data already taken, or both...?

Reply:

Cf comment 12.

Comment 077 by PrebenGrosbol

p.32 6.3.4-R6
'must be passed or ...' Same comment as for 6.3.3-R1

Reply:

Agreed.

Comment 078 by WimBrouw

p32
6.3.4-R5: drop the 'and aperture efficiency'
Add 'squint' to the list; or just make it 'generic beam parameters like ...'
Should this also be a plug-in? I assume this is not done on a daily basis.

Reply:

Requirement rephrased.

Comment 079 by WimBrouw

p32
R6: drop

Reply:

Requirement rephrased.

Comment 080 by JoeSchwarz

p. 33, 6.4-R2 This seems too vague to me. 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? It's entirely unclear how this data is to be used, e.g., whether we're talking about letting the PI play operator from his/her institute in Europe or America or just check on what an image looks like from time to time...

Reply:

This was introduced last year after the previous review. The way it should work is explained in 4.0-R9 and Use Case 4.5.2.

Comment 081 by PrebenGrosbol

p.33 6.4-R1
'shall be activated automatically after each ...' That is after each observation which may be too much. I would weaken the statement and not make it mandatory but configurable.

Reply:

Requirement rephrased.

Comment 082 by PrebenGrosbol

p.33 6.4-R2
'..., via the Internet' It sounds nice but it may give a bandwidth problem if lots of people tries to get Quick Look data like images.

Reply:

Cf comment 80.

Comment 083 by WimBrouw

p33:

6.4.1-R3.4: integrated? Since the session will contain 'simple scans' in frequency and/or position (mosaicing..) difficult to define exactly here. Just say 'indication of above over session' or so

Reply:

Rephrased.

r3.5: noise per pointing?

Reply:

Added R3.6: for mosaics, it should be possible to monitor all quantities (R3.1 to R3.5) per pointing center

Comment 084 by WimBrouw

p33

3.6.4.2/3: See it as an indicator of what should be done at the minimum. A good list. Maybe add to R3: 'TAKING INTO ACCOUNT FLAGGING'

Reply:

It is implicit in the definition of 'flagging' that all data affected are not taken into account in subsequent operations, unless specifically mentioned.

Comment 085 by JoeSchwarz

p. 34, 6.4.2-R4 "Mosaic and self-calibration projects shall be supported." We need much more information than this to understand this requirement, and hence to know whether we have fulfilled it! What kind of Quick Look Operations are appropriate here?

Reply:

Requirement rephrased.

Comment 086 by MasatoshiOhishi

p. 34 3.6.4.3 Data Processing: Single Dish Data

In 6.4.3-R1.1, "on/off " could be replaced by "position switch".

Reply:

Agreed.

Comment 087 by TimCornwell

Page 34, 6.4.2-R3:

The mention of the Fourier transform is too prescriptive, and is unnecessary. It is too prescriptive because it is entirely possible that we might end up using some efficient linear algebra technique to make images quickly. It would be better to require something like along the lines that any image be available in a time comparable to a typical observing block.

Reply:

Agreed. This requirement is rewritten accordingly.

Comment 088 by TimCornwell

Page 34, 6.4.2-R4:

More explanation is needed: does the Quick Look pipeline just process an observing block or all relevant observing blocks? E.g. in a mosaicing experiment, does the QL pipeline just image the last patch of the sky observed? This is also a problem for self-calibration.

Reply:

The QL pipeline deals with all data acquired during the current session. For a mosaic, it should probably use a quick-and-dirty method to produce an image, the best processing being deferred to the Science Pipeline and/or to the off-line stage.

Comment 089 by TimCornwell

Page 34, 6.4.2-R5.

This is too prescriptive and probably wrong: one is better off doing this in the visibility plane. I'd make this a best practice requirement.

Reply:

Agreed. See comment 91.

Comment 090 by WimBrouw

P34

6.4.2-R4: mosaic yes; self-calibration not as stated (see R3; and many earlier remarks about 'no or limited'. Should be limited to make it a non-bottleneck

Reply:

Agreed.

Comment 091 by WimBrouw

P34

R5: Compare to 'clean-beam'?? I do not get this at all. Maybe state: For ...the pipeline shall use the data to produce an estimate of the seeing.' (and pointing??)

Reply:

The idea was to check the size of the source with that of the clean beam, to check the seeing effects. This requirement has been rephrased to be less prescriptive: the pipeline shall be able to make any valuable test on the data taken on a point-like source.

Comment 092 by WimBrouw

p34

6.3.4-R1: a question: no statement about 'supported modes' is made for synthesis. Switching, OTF etc could be there as well. Should be added somewhere?

Reply:

There are actually less observing modes for aperture synthesis than with a single-dish antenna: single-field imaging, mosaicing, OTF mosaicing, snapshots. All modes can actually be calibrated in a similar way, the imaging processing being different. Requirements 6.4.2-R3 and R4 have been modified to make it more explicit.

Comment 093 by JoeSchwarz

p. 36, 6.5-R1, R2 The term "session" is used too imprecisely here. Isn't it possible that some data preceding and following a session will be relevant? If, for example, a baseline calibration was done by the project that was executing 5 minutes before the current one, do we really want to repeat it? I imagine that there is other calibration data coming from outside the "session" that could be useful, too.

Reply:

Several calibrations are to be done by the ALMA staff, and results used by all subsequent projects. This includes baseline, pointing model, primary beam measurements, etc. The results of such observations are to be used by the real-time control system, to update antenna parameters.

For 6.5-R1.1/6.5-R1.2, 6.5-R2, only very recent data is probably usable (same receiver tuning needed).

I would also change "shall find in the Archive all data..." to "shall use all data...". Where this data comes from (whether it's in the Archive, or cached in memory, or whatever) is irrelevant from a requirements point of view.

Reply:

Agreed.

Comment 094 by MasatoshiOhishi

p. 36 3.6.5.2 Single Dish Data

In 6.5.2-R1.1, "on/off " could be replaced by "position switch".

Reply:

Agreed.

Comment 095 by TimCornwell

Page 36, 6.5-R2:

Can and should the Science Calibration Pipeline make use of historical observations which are not part of the project?

Reply:

Yes: I think all observations of known sources identified as calibration observations should be usable by all projects. In case a particular project is observing a famous calibrator e.g. NRA0150 to search for specific lines at high frequency resolution, then these observations should be identified as target observations and will not be considered by the system as calibrations. I think the main interest is for flux monitoring of calibrators, anyway, and that should be handled by an observatory project using all calibrations, not by each project directly.

Comment 096 by WimBrouw

p36

6.5.1-R1: The way the bandpass and time-variations are given here does not indicate that the requirement could be for multiple bandpasses (depending on type of session) and time-frequency variations.

Reply:

Agreed there may be several bandpass calibrations in a session.

Comment 097 by WimBrouw

p36

R2: should the pipeline 'observe' such a source if none available??

Reply:

No. It is the responsibility of ALMA to make sure that the observing modes are consistent: The observing procedure should be intelligent enough to ensure that a project is not done without checking that e.g. flux and a bandpass calibration are performed. The Pipeline will not go back and observe.

Comment 098 by WimBrouw

p36

R3: why not drop R3.1 and make it just amplitude and phase corrected for the appropriate frequency ('scaling' could be too simple depending on how the delay is done as a mixture of time-delay and phase rotation??

Reply:

Rephrased.

Comment 099 by JoeSchwarz

p. 38, 6.6-R1 "subproject" is not defined anywhere in this document.

Reply:

Changed to 'project'.

Comment 100 by JoeSchwarz

p. 38, 6.6-R4 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 (because, for example, it requires different antenna configurations), why can't he/she wait a few more days for the delivery of a DVD? I think that this decision should be made in view of a) how urgent the project is (a basis for coordinated or follow-up observations?); b) how much data is involved (Alma will produce 180 Tb/year, so it's not impossible that we might occasionally have datasets that are significant fractions of a Terabyte); and c) what the real capacity, cost and reliability of the Internet is in the Alma epoch.

Reply:

Agreed. We have rephrased the requirement, the data should be in the archive and accessible to the PI (in fact I do not think we have required that the data are 'delivered' to the PI).

Comment 101 by JoeSchwarz

p. 38, 6.6.1-R1, R2 "...find in the Archive..." Again, where it comes from isn't relevant.

Reply:

Agreed.

Comment 102 by JoeSchwarz

p. 38, 6.6.1-R3, R4 If the flux scales are different, what is to be done? What are the possible consequences of "direct comparison of the redundant data"?

Reply:

In both cases, if the data are not compatible, the quality flag should be down.

Comment 103 by JoeSchwarz

p. 38, 6.6.1-R7 I suggest that having "several [deconvolution] algorithms running in parallel" is better done as an offline, rather than an online task. If it *must* be done as part of the production pipeline processing, I don't see how it can be met as a "priority 1" requirement, which, as Page 12 of this document defines it, means:

"Must be there for Interim Science period, when the system is commissioned to produce meaningful science results."

Reply:

Requirement rephrased: having a deconvolution algorithm available is in Priority 1. Having several algorithms running in parallel is in priority 3.

Comment 104 by MasatoshiOhishi

p. 38 3.6.6.1 Interferometric Data

In 6.6.1-R7, it would be worthy to add a new CLEAN algorithm, the Wavelet-CLEAN, developed by Japanese group. And it would be very important how we judge the image quality. Which is the best ???

Reply:

We should add this, shouldn't we? but not be restrictive to a fixed list of algorithms.

Comment 105 by PrebenGrosbol

p.38 6.6-R4

'..., via the Internet' Same comment as for 6.4-R2

Reply:

see comment 100.

Comment 106 by TimCornwell

Page 38, 3.6.6 Science Imaging Operations

There is a potential problem in all pipeline processed images: the provenance of each scientific result must be knowable and straightforward. If a virtual observatory is to make use of ALMA results, there must be an ALMA standard product that was produced in some standard way without strange choices for e.g. cellsize, field of view made by the observer. By this logic, one is forced to produce at least two results: the "standard product", and that required by the observer. In many cases, the observer may just ask for the standard product. This whole aspect of the scientific imaging pipeline must be clarified. It affects a large number of the requirements following. My specific recommendation would be that two products are produced: the standard product defined by ALMA, and the observer's product, defined by the observer.

Reply:

The idea of having two products is reasonable, provided the non-standard products are not required too often for large images (in order not to increase the archive size needlessly. Note that I have no idea how to define the standard product in terms of frequency sampling: using the whole available correlator output at all times would boost up the data rate! Here I think we have to trust the observer, as well as with the mapped area. So what remains are the cell size and the mapping algorithm. We have control over those as we provide them in the pipeline.

My guess is that we should take your option whenever the user-provided parameters (e.g. cell size, algorithm) are not regarded as 'safe'

(e.g. may degrade the fidelity more than we like) by the pipeline.

Another question that must be resolved is whether results are processed only before insertion into the archive or also on exit from the archive (e.g. triggered if best practices have changed). This is a hard question to answer. This is discussed in requirements below but I think it needs to be thought through a bit more.

Reply:

I think we should have this possibility. This is mentioned in the Introduction and in the Archive section.

Finally, is the observer free to use the pipeline repeatedly or is the processing limited to that specified in the observing setup? If the former, how is time allocated? I think the latter is therefore preferred.

Reply:

These are operational issues. The idea was that at regional centers

the pipeline may be run to reprocess data to the current standards of data reduction. Experience seems to show that reprocessing data one year after observations can be done at a fraction of the time (and cost) of the first processing. But I agree this may be a problem for large projects.

Comment 107 by TimCornwell

Page 38, 6.6-R2

In many cases, the final product could be a linear mosaic instead of a deconvolved image. I'd therefore remove the second sentence.

Reply:

In all cases, a deconvolution will be applied to the data, so we prefer keeping the current phrasing.

Comment 108 by TimCornwell

Page 38, 6.6-R3

Can relevant non-proprietary data from other projects be included?

Reply:

This is a policy decision obviously. My position would be yes, provided that the availability of such data was checked before proposing; then the reviewers and/or project management may decide...

Comment 109 by TimCornwell

Page 38, 6.6.1-R3

Phase centers and polarization frames must also be checked.

Reply:

This is included in the general term "instrument setup and properties".

Comment 110 by TimCornwell

Page 38, 6.6.1-R6, 6.6.1-R7.

If ALMA works as well as expected, the extra images requested here will be unnecessary: for example, the various weightings should be less divergent than for the VLA. Similarly for the deconvolved images. Also there is a combinatorial explosion possible (weighting x deconvolution x other parameters). I'd recommend that the Standard Product be just one image.

Reply:

Yes, but the pipeline should have the possibility, if explicitly required, to use different weighting schemes.

Comment 111 by WimBrouw

p38

6.6.1-R5: There is no 'continuum measurement'

Reply:

Requirement rephrased: the Pipeline shall extract visibilities of a PSEUDO-continuum, corresponding to the sum of the spectral channels.

Comment 112 by WimBrouw

p38

R2: 'compatible'?

Reply:

Yes

Comment 113 by WimBrouw

p38

R5: 'appropriate'? I would think that the data is used to produce the required image cube; not the other way around.

Reply:

Yes

Comment 114 by TimCornwell

Page 39, 6.6.1-R9

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!

Reply:

Requirement rephrased.

Comment 115 by WimBrouw

p39

6.6.1-R7.3: Add: 'model fitting and data subtraction'

Reply:

Yes

Comment 116 by WimBrouw

p39

R8: do not specify both the domains. Why? I would think that image-plane subtraction is a non-option in general.

Reply:

Yes

2 Comments on Archive Section (3.7) and replies*prepared by Ken Tatematsu and Kouichiro Nakanisi**old means item number in reviewed No. 11 doc**new means item number in new draft***Comment 007 by TimCornwell**

- I think that the distinction between science and technical archives is not worth making, and would probably lead to problems.

Reply:

related to 118, 145,

In the original No. 11 document, the technical data seems to be only in the header of (observational) data.

The draft authors thought that we must store the technical data both during observation

and when telescope has no observation.

OK, let us use "observational data" and "technical data" rather than two functionality of the archive, because it may be misleading.

Comment 009 by WimBrouw

Comments on Pipeline part of ALMA SSR and Use cases -- version 4

Question (in principle for Use cases): will the measured WVR be archived? If not, calibration based on WVR should not be included in off-line package.

Reply:

[see old3.7.2-R3=new3.7.2-R3](#)

Comment 117 by JoeSchwarz

p. 40, Section 3.7 in general

It is likely to be very difficult to construct a satisfactory (User-) Archive without more information about its intended use. It seems to me that statements like "all data taken by the array is archived" are somewhat irresponsible. Certainly, we don't want to lose anything important, but this shouldn't be used as an excuse to not think about what should be in and what should be left out. It's not just that storage space is wasted, but that packing the Archive with data that's not needed makes it more difficult to manage and organize the data that *is* needed.

Reply:

[related to 119, 122, 155, 156](#)
[comment 155 is accepted](#)

Comment 118 by JoeSchwarz

p. 40, 7.1-R2 The distinction between "observational" and "technical" archives is never exploited and is often blurred (scripts, for example, seem to be included in both). What is needed is specification of what uses will be made of what kind of data. Then the Archive designers can decide where best to put it to facilitate these uses.

Reply:

[related to 009, 145](#)
[see reply to 009](#)

Comment 119 by JoeSchwarz

p. 40, 7.2-R2 "The observational archive shall also include as header data... technical data..." So, again, why have we defined a "technical" archive that is to hold "all technical data"?

Reply:

[related to 122, 155, 156](#)
[comment 155 is accepted](#)

Comment 120 by JoeSchwarz

p. 40, 7.2-R5 "...extract the database information for efficient data search from the header..." What does this mean? That certain information should be indexed? Why not specify what kind of performance is desired (i.e., what "efficient data search" means) and let the Archive designers worry about how to get it.

Reply:

related to 129,134

see replay to 134

Comment 121 by JohnBenson

p.40

7.0-R1 I think you should also archive the calibrated/flagged data from the pipeline. This is probably what the observers will want distributed to them anyway..

Reply:

We guess 7.0-R1 should read 7.2-R1.

Is "images produced by the pipeline" not OK?

The pipeline has automated flagging (From old3.6.2-R3).

Comment 122 by JohnBenson

p. 40

7.0-R2 In order for dynamic scheduling and automated pipeline processing to work, the system needs to have a quantitative, parameterized description of the observers scientific goals. Things like sensitivity limits, image fidelity requirements, maybe spectral range and resolution in km/sec...

Reply:

related to 119, 155, 156

We guess 7.0-R2 should read 7.2-R2.

We think these are in the first item "all user (observer) input".

comment 155 is accepted

Comment 123 by MasatoshiOhishi

p. 40 3.7.2 Observational Archive

In 7.2-R2, item 4, wrong reference to 1.3-R2 ? It might refer to 2.3-R2 in page 17.

Reply:

It refers to old3.7.3-R2, thanks.

However, the revision made this reference unnecessary...

Comment 124 by PrebenGrosbol

p.40 7.2-R1

'... shall include raw data, header information, ...' All these items should be in the archive but I was missing things like catalogs of line data or calibration sources. Where are they going to be?

Reply:

related to 184
accepted (new3.7.2-R2)

Comment 125 by TadafumiTakata

p40 7.1-R1

Comment

Archive system should enable astronomers and engineers to know the status of data handling. It is like "supervisor of data handling".

On managing side, it is very useful to know where the data is at that time such as in the way of pipeline processing, archiving, or in the way to RSC or so, especially in the trouble around data handling.

The most important role of archive system is let users know everything about all data.

In the distributed database environment such as ALMA, which has several copies of database in each RSC and so on, tracing the problem in data handling is very complicated work and this supervising function

(like FEDEX managing system,,,) may be useful in various sites of ALMA.

On user side, such as astronomer(observer and/or support astronomer, observation operator) may want to know the status of their observational data such as what step their data are in pipeline processing and so on. For astronomical requirement, the trace of dataset processing may be better for users.

Reply:

partially accepted

will show the data list, when requested.

The pipeline status will not be necessary for ALMA archive, because the archive user's concern is just whether the data is ready for reduction or not.

Comment 126 by TadafumiTakata

p40 7.2-R4

Comment

The submission of offline reduced data by users should be performed using user-friendly GUI and submission process should have the function to make a link from submitted data to the original raw data or dataset for effective use of these data by archive users or so.

Reply:

GUI described in old3.7.4-R1=new3.7.3-R1 will guarantee the first point.
The latter part is accepted

Comment 127 by TimCornwell

Page 40, 7.2-R2

Throughout this section, the phrase "header information" is used. This

seems to indicate to me that every item retrieved must have the extensive information attached. In the first sentence here, I'd just remove "as header data".

Reply:

accepted

Comment 128 by TimCornwell

Page 40, 7.2-R3

I cannot think of any reason why one would want to let the user make an irreversible choice like this. I'd remove the last sentence "This may be overridden."

Reply:

accepted

Comment 129 by TimCornwell

Page 40, 7.2-R5

I have no idea what this means!

Reply:

see reply to 134

Comment 130 by WimBrouw

p40

3.7.1: I would suggest to have the V0 requirements R1 and R3 (7.5) in the introduction. They are essential requirements from the science point of view

Reply:

related to 185

We think that V0 related items should be located in a subsession, because it treats some specific function.

Comment 131 by WimBrouw

p40

7.1-R2.1 : add: Access and interface to the two archives should be compatible

Reply:

see reply to 007, 167

R2.2 : add: Searching in the archive should be of $O(1)$, or $O(\ln N)$ at most.

Reply:

We think that old3.7.1-R1=new3.7.1-R1 includes this.
More clearly we state this in old3.7.4-R2=new3.7.3-R3.

Comment 132 by WimBrouw

p40

7.2-R3: I think it is wrong in principle to let the user (I suppose the observer here) decide what should be archived. The value of an archive lies in having information of which it is a priori unknown if it will ever be used. Either always archive both (preferred IMO) or let the ALMA operations at some stage decide which one will be archived from then on

Reply:

accepted

Comment 133 by WimBrouw

p40

7.2-R4: This issues raises the questions of archive management (which should maybe be raised in use cases) and of 'super-headers' ';linking' etc. In cases of final products and/or papers would it be advantageous to have the headers of observations used point to these outcomes (and the outcome of course point to the observations). Both would probably best served by having a 'supra' or 'virtual' header describimg the sum of observations used; and the results (with, if at all feasable a pointer from low-level to high level (or linked-list)). Note that one observation can be part of many super headers.
Probably this requirement is one for the introduction.

Reply:

This will be treated in the section of V0 and interface to it.

Comment 134 by WimBrouw

p40

7.2-R.5: Like I mentioned above, I think efficiency should be stated as a general requirement, and not give a 'solution'. I could easily imagine that there will be sub-headers (e.g. log or linkage to others, or because knowledge increases) which should be searchable. Drop this one, and replace it with:
"All information describing the observation should be accessible through the header."
This ensures that the archive will as coherent as possible (and is also updatable easily at a later stage).

Reply:

accepted

Comment 135 by WimBrouw

p40

7.2-R6: why not:

All information shall be archived in SOC/OSF as soon as it is available in coherent units.

This will enable headers; logs; rawdata in pieces; pipeline parts (like individual channel images) be off-loaded as soon as possible

Reply:

Related to 171

Comment 136 by JoeSchwarz

p. 41, 7.2-R8 Shouldn't this be the responsibility of the Regional Centers? Why should there be an *additional* access point?

Reply:

We think that the principal archives represent those at SOC, and RSC with physical archive.

Note that the recent ASAC report says that RSC might choose just a link to SOC.

Comment 137 by JoeSchwarz

p. 41, 7.2-R9 This requirement can be dropped if each Regional Center has a full copy of the Archive. Backup of this amount of data is a *very* expensive exercise.

Reply:

related to 142, 147

At any rate, we may need some kind of backup
(another RSC or shadow archive or tape...)

Comment 138 by JoeSchwarz

p. 41, 7.2-R11 Who decides what the "goal" of a scan is? Can't it serve multiple purposes? If an "expert" writes a series of low-level commands to define his/her observing procedure, how can the system figure out what his/her intention was?

Reply:

We think that it must be defined in OT, and then the info is to be stored in the archive.

In case of low-level commands, how about putting them for target observation?

Comment 139 by JoeSchwarz

p. 41, 7.2-R13 If we should sometimes store images and generate them on-the-fly at other times, then I conclude that uniformity is not a requirement for the Alma Archive. This could pose a problem for survey-type research (i.e., to make sure that you're getting a uniform sample, you'd have to reprocess all the images that you want to include). Is this really what's wanted?

Reply:

related to 140, 152, 153

When we have new calibration data or method, the stored image will be recalibrated (old3.7.2-R14=new3.7.2-R13.3)

Comment 140 by JoeSchwarz

p. 41, 7.2-R14.2 Does this mean that, for example, an e-mail should be sent to anyone who has ever used the Archive every time calibration procedures change? This sounds a little like spamming.

Reply:

accepted

When necessary, data will be recalibrated automatically.
related to 139, 152, 153

Comment 141 by JohnBenson

p. 41

7.0-R8 I think the archive should be accessible through web-tool GUI's on the internet. Essentially all information in the archive catalog tables (your header data I think) should be accessible to any qualified user.

Reply:

acceptd

Incorporated in the GUI section old3.7.4=new3.7.3

Our goal with the NRAO E2E Archive is to build web-tools that allow a user a wide set of queries, and allow selection and FTP downloading of archived data. I think the FTP downloading will be very popular for a substantial fraction of observing programs. The rest will have to be distributed on some recording media.

Reply:

see old3.7.4-R16=new3.7.3-R17

Comment 142 by MasatoshiOhishi

p. 41 3.7.2 Observational Archive

In 7.2-R9, only one backup for the archive ? It would be useful to backup the principle archive in ALL RSCs.

Reply:

related to 137, 147

see reply to 137

Comment 143 by PrebenGrosbol

p.41 7.2-R13.4

'Image must always be archived if the pipeline cannot ...' This requirement just re-states a special case of 7.2-R13.3.

Reply:

acceptd, delete this

Comment 144 by PrebenGrosbol

p.41 7.2-R14.1

'... and provide the most up-to-date calibration.' It sounds like only the most recent calibrations are in the archive. I would think that all calibrations are there but you by default only get the latest.

Reply:

acceptd

Comment 145 by TimCornwell

Page 41, 3.7.3 Technical archive

The concept of a technical archive is troublesome. There are different roles (observer, engineer, operator) that access the archive but I do not think that this should mean that there are different archives. I would rather see one archive that can be filled to analysis packages in different ways: e.g. some tables to engineers, most information but perhaps sub-sampled to analysis programs. In the AIPS++ MeasurementSet, we have subtables for environmental data (the WEATHER subtable) in the expectation that the observer might chose to flag all data where WINDVELOCITY>20m/s. Similarly the observer might wish to see the operator log book. Hence the trend that I see is to move away from separate archives but allow filling programs to chose to fill different information from the archive in a context-dependent way.

Reply:

related to 007, 118
see reply to 007

Comment 146 by WimBrouw

p41
7.2-R7: There may be several SHADOW archives ..

Reply:

accepted

Comment 147 by WimBrouw

p41
7.2-R9: One shadow archive shall act as a backup, and be shadowed on a
continuous and complete basis
Note: the shadowing could be done through a tape or other medium, not
necessarily net.

Reply:

related to 137, 142
will not specify implementation (mirror or tape).

Comment 148 by WimBrouw

p41
7.2-R10: does R6 not already imply this (i.e. how can you archive data if you
have no access to archive?)

Reply:

see reply to 171

Comment 149 by WimBrouw

p41
7.2-R11: should be 2-way. I.e. header should know what the data (or dat
piece) represents (cf R5 (either old or proposed from)

Reply:

?

Comment 150 by WimBrouw

p41

7.2-R12: ceratinly at the start of the ALMA observations (and even much later) it will be completely unknown which technical/environment data could/should be used at some (maybe future) time to improve the quality of the reduced data. Limiting it at this stage to 'which is necessary to make off-line analysis' and 'if not present in header' is too constricting.

Why not:

'The archive shall provide, in the observation header, the appropriate links to all technical data available for the observation period. This link is in addition to the set of technical data that is provided in the header.

Reply:

accepted

Comment 151 by WimBrouw

p41

7.2--R13.5: Add: On-the-fly re-imaging should be available for saved images as well in special circumstances

This to be able to redo bad calibration; or to compare with a newer observation done with a different calibration scheme.

Reply:

We think that it is guaranteed from other requirements, isn't it?

Comment 152 by WimBrouw

p41

7.2-R14: I sthis talking about calibration 'DATA' or 'PROCEDURES' or both?
both

Does transparently mean that all the maps in the archive are invalidated and/or automatically redone?

Or just a message/code to any user of a stored map from before the change?

Reply:

related to 139,140,153
see reply to 139

Comment 153 by WimBrouw

p41

7.2-R14.3 '... which should always use the 'standard' ...' is undefined. The standard will be different for images that are retrieved from stored to images done at the same time but re-imaged OTF. I suggest that if a stored imaged that is reflagged is retrieved, it will be re-imaged always OTF.

Reply:

related to 139, 140, 152

see replay to 139

Comment 154 by WimBrouw

p41

7.3-R1: replace 'recorded' by 'measured'. Changed the part after ',' by: Each item shall contain a time-stamp.

The reason is that you should not prescribe the archive organisation: a more hierarchal or other structure could be better than just a single threaded time series.

(Even by requesting a time-stamp you request certain hardware characteristics of the information gathering devices).

Reply:

accepted

Comment 155 by WimBrouw

p41

7.3-R2: Always dangerous to give an inclusive list as requirements. At the least say:

- all measured environmental data
- the water vapor radiometric raw data (at ~1s timescale)
- all monitored data

I should exclude the 'derived' pathlength. This is model dependent (and hence will vary with time), but also it is cheaper to recalculate (processing is always faster than reading data up to quite some processing; and it will save storage space).

Reply:

related to 117, 119, 122, 156

accepted

Comment 156 by JoeSchwarz

p. 42, 7.3-R3 The "high- and low-level scripts" that we're asked to save in the Technical Archive were already saved in the Observational Archive in 7.2-R2. This reinforces the need to be explicit about what is to be *done* with the data, not *where* it is to go.

Asking for the "monitor data" is pretty vague and open-ended. It would be good to think about what's really needed here.

Reply:

related to 119, 122,155
 e.g. the record of "go to E1=45 without observation"

The examples of monitor data are:

- SIS bias voltage and current
- Total power from IF port
- temperatures inside the cryostat
- PLL error voltage

Comment 157 by JoeSchwarz

p. 42, 7.3-R3 and R4 Things like electronic log books and records of manual operations should be available in the Archive, but it should be clear that it's not an Archival task to provide the user interfaces to enter this data.

Reply:

We think that it may be a task in the archiving system UI.

Comment 158 by JoeSchwarz

p. 42, 7.4-R1 says that the Archive Search Tool should be a GUI, 7.4-R2 states that the Data Extractor Tool should use it as a front end, 7.4-R9 says that 'the Data Extractor Tool shall use the Search Tool'.... This is a clear contradiction! Which tool is using which other tool, shall a programmatic tool (Data Extractor Tool) use a GUI (the Archive Search Tool)?

Reply:

accepted

in old3.7.4-R9=new3.7.3-R10

" the Data Extractor Tool shall receive the information from the Search Tool"

Comment 159 by JoeSchwarz

p. 42, 7.4-R5 "Two interfaces" are asked for, but the differences between them are never described.

Reply:

related to 145

see reply to 007, 167

We think that we need separate interfaces for separate purposes. Difference would be obvious from the purpose.

Comment 160 by JohnBenson

p.42

7.0-R10 You might add 'project' and 'observer' to your list of search criteria. 'Molecular transition' is a good idea, I hadn't thought of that one for the E2E Archive, I'll use it.

Reply:

We think

- Project name
 - Name of Principal Investigator
- correspond to them.

Comment 161 by PrebenGrosbol

p.42 7.3-R3

'The archive shall record all high- and low-level scripts' It seems not reasonable to request the archive to record scripts, it should rather store them. It must be the Dispatcher/Sequence which records the scripts in the archive.

Reply:

accepted

Comment 162 by TadafumiTakata

p42 7.4-R6

Comment

Especially in technical archive, it is very important to provide the function of data searching using any kind of keyword and header information flexibly. (R6 may be very essential especially during developing and first light phases.)

Reply:

We think that "Any other header info" corresponds to this.

Comment 163 by TimCornwell

Page 42, 7.4-R1

A CLI must be available to interrogate the archive from scripts. In addition, one will also want a web service equivalent.

Reply:

accepted

related to old3.7.4-R14=new3.7.3-R15

Comment 164 by WimBrouw

p42

7.3-R4: make it priority 1: especially at the start of operations the notes can be very helpful in disentangling any problem or error.

Reply:

accepted

Comment 165 by WimBrouw

p42

7.4-R2: ... searching the OBSERVATION database ...

Reply:

accepted

Comment 166 by WimBrouw

p42

7.4-R4: priority 2 (make the cookbook then priority 1): use feedback to finalise help

Reply:

accepted

Comment 167 by WimBrouw

p42

7.4-R5: Do you mean:

The AST shall have two interfaces: one mainly for astronomical product production; one mainly for technician use (e.g. for quality monitoring and error tracing)

I believe that data from both archives can (and should be) used for either astronomical and technical purposes. The interface for both purposes should be different (certainly), not the underlying available data.

Reply:

accepted

Comment 168 by WimBrouw

p42

7.4-R6: 'The search criteria shall include all the information in the observation headers, including any information pointed to in sub-headers. Search criteria based on combination of search fields should be possible (e.g. time * bandwidth) They shall include (but not be limited to) e.g.:

- your list; but:

is integration time the total or the per sample;

Reply:

integration per image pixel, I think

why not have a product of bandwidth and time (for continuum mostly)
there are coupled items: configuration and resolution; frequency and
resoltion, ...

Reply:

accepted

Comment 169 by WimBrouw

p42

7.4-R6.1: Add: The user interface should be able to understand an SQL-type
language with expressions between fields.
(this takes also care, and should be combined with R.7)

Reply:

see [old3.7.4-R14=new3.7.3-R15](#) revision

Comment 170 by JoeSchwarz

p. 43, 7.4-R8 The intent of showing "query statements" isn't clear. We don't
know what database technology we will use for the Archive; I assume the user
just wants to know what the search criteria was and be able to modify them.

Reply:

accepted

Comment 171 by JoeSchwarz

p. 43, 7.4-R12: This is impossible to provide for a general archive user.
We must
allow for some (possibly short) delay. The Scheduling Process and ALMA
operations must have top priority at this stage. If we still think about
something like a 'Fast Store Archive' and a separate general Archive, then
the upload of the data to the general archive will have to be asynchronous
and might be delayed in case of peak load to the fast store.

Reply:

related to 135

"immediately" is replaced by "as soon as practical"

Comment 172 by JoeSchwarz

p. 43, 7.4-R13 How can a Data Extractor provide links? Or is the Data Extractor really a GUI?

Reply:

Date Extractor should read Search Tool.
Thanks!

Comment 173 by MasatoshiOhishi

p. 43 3.7.4 User Interface
In relation with 7.4-R13, it would be useful to have a hyperlink with published papers e.g., ADS if available, that used the relevant data.

Reply:

accepted
although the term "hyperlink" may be modified (see 181).

Comment 174 by PrebenGrosbol

p.43 7.4-R14
'... request to get the file like' The requirement should be more specific and not just give one example. Also this requirement seems to contradict 7.4-R9 which states that the Search Tool shall be used.

Reply:

see reply to 172, 182

Comment 175 by RayPlante

p. 43, 3.7.4, 7.4-R13
Wording seems a little funny. I think you mean that the Tool, when displaying project/data product descriptions, should be able to include hyperlinks that retrieve related images or catalog data from external archives.

Reply:

see reply to 172

Comment 176 by RayPlante

p. 43, 3.7.4, 7.4-R14:

I didn't quite understand what this meant. Does it mean...

- a. the Tool should respond to a url-encoded HST archive search query? (I don't think so.)
- b. that the Tool accept URL-encoded search queries for retrieving data? This implies that the tool is a web service (as opposed to a client application that connects to the archive via the web).

Personally, I don't think it is sensible to retrieve big data products via a search query as it's too easy for the query to return much more data (or not enough) than you want. Instead, associate each data product with a unique identifier, which is then retrievable via a unique URL. Search queries can then return these URLs along with other metadata; the Extractor (or user) can then select what data should actually be downloaded. This is the model used with the BIMA archive. The client application DaRT allows the user to download a list of URLs all in one shot.

Reply:

see reply to 172 and 182
"search requests" added

Comment 177 by WimBrouw

p43

7.4-R10: Is meant as: 'A preview image of the image produced from the selected data shall be made available before transfer of the final image-cube'?

Reply:

yes

I think this requirement is not meant that way, but what is meant to produce from the selected data a small image. Small in field? Integrated over spectrum? Low resolution? Taking only every 10th datapoint? Whatever way you do it, for it to give any indication of correctness, a full calibration and imaging must be done.

Reply:

yes

Maybe it would be better to change this to something like:

Before transmitting the selected data, an image of the distribution of the datapoints in the Fourier domain will be transmitted (or maybe a PTF) at some centre frequency.

Reply:

Preview image is to choose what should be transmitted.

Comment 178 by WimBrouw

p43

7.4-R10.1: cannot be done: the data selection will not correspond with pipeline result. Only way to do this is to add all the individual pipeline results (will they always be scaled the same?) images based on the data selected.

Reply:

If integrated over relevant velocity and pixels are small, OTF would be fast enough, we think.

Comment 179 by WimBrouw

p43

7.4-R10.2: same m.m.

Reply:

(idem)

Comment 180 by WimBrouw

p43

7.4-R11: is that not inherent in 7.2-R6?

Reply:

R6 is just search criteria only.

Comment 181 by WimBrouw

p43

7.4-R13: replace 'hyperlink' with 'persistent link information' (hyperlinks have the tendency to become invalid; by using more modern ideas (e.g. XML links) this can be largely overcome: leave it to the archive designers to come with solution)

Reply:

accepted

Comment 182 by WimBrouw

p43

7.4-R14: should not be as such in SSR. More:

The DET should be able to accept (properly verified) web-based extraction requests

Reply:

[accepted](#)**Comment 183 by WimBrouw**

p43

7.4-R15: The DET must be invocable from the Offline ..

(I suppose you do not want to limit it to Offline package only: must be stand-alone as well (and R14))

Reply:

[accepted](#)[oldR14=newR15 changed](#)**Comment 184 by JoeSchwarz**

p. 44, 7.5-R2.1 I believe there was an ASAC decision to exclude source extraction as part of the Alma project responsibility. If this is right, what are the "ALMA Catalogues", and who will produce them?

Reply:

[related to 124](#)[old3.7.5-R2.1 deleted](#)

We may contain catalogues made outside the ALMA project, but can be as RSC task.

Comment 185 by JoeSchwarz

p. 44, 7.5-R3 This requirement says basically, "We don't know what the VO requirements are, but you must meet them, and moreover as Priority 1 [when Interim Science Ops begin]." This will be an exceptionally hard requirement to meet.

Reply:

[related 130](#)[see modification](#)

Comment 186 by MasatoshiOhishi

p. 44 3.7.5 Relationship with the VO Projects

As one of leading persons for the Japanese Virtual Observatory Project, I would very appreciate to see this section. Important points are to guarantee data quality including in providing its reliable information, and to provide network-transparent interface to connect to each VO via, for example, the globus tool kit.

Reply:

[related to 188](#)

Comment 187 by PrebenGrosbol

p.44 7.4-R19

'... secure access to proprietary data' Security is good and needed but I would also have expected some more general requirements on access like that all users of the archive should be identified by user-id or something like that.

Reply:

[accepted](#)

Comment 188 by TimCornwell

Page 44, 3.7.5 Relationship with Virtual Observatory Projects

There are several blank checks being signed here! There is no definition yet of what it means to meet requirements for VO access so the project cannot take on an obligation to support such access. I would remove this section (I am in favor of VOs but not of signing blank checks).

Reply:

[related to 185,186](#)

[see reply to 185](#)

Comment 189 by WimBrouw

p44

7.4-R17: If an archive user requests a DISK file... in A user accessible directory. The user will be informed about the estimated transfer time. An email message can be requested at the end of the transfer.

Reply:

[accepted](#)

Comment 190 by WimBrouw

p44

7.4-R19: is DET or AST meant her (R9 says that the DET uses the AST). Or is it meant that login only necessary to get proprietary data?

Reply:

Data Extractor Tool should read Archive Search Tool.
Thanks!

Comment 191 by WimBrouw

p44

7.5-R1/R3: see earlier

Reply:

Comment 192 by WimBrouw

p44

7.5-R2: Whatever ... to provide:
- ALMA catalogs (i.e. archived images)

Reply:

related to 124, 184

- image quality information for archived images
- data quality information for selected observations

Reply:

accepted

Comment 195 by WimBrouw

p77

4.6.2: The goal There are proprietary ... always available for everyone.

This is stated to sloppy in view of earlier comments in the archiving part that all search information should be available in header. This could e.g. be the max/min flux (as it should be I think), integrated flux; SNR; binning period for variable phenomna; source list coordinates etc which could be indicators of the astronomical results. Maybe 'header' should be replaced by 'pre-observational header information' or some other restriction

2. Does user provides all of these? some of them; as an example?
exception Course 1: PART OF requested..

Reply:

accepted (in Search Archived Data UC).

Last issue mentioned (pipeline output): is already completely defined in the requirements

Reply:

The modification (including above) made by K. Tatematsu and K. Nakanishi for Use Cases for archiving was not incorporated in the the new No. 11 for review, by accident in the editorial process. We incorporate all revisions (we made before review and after review) in the new draft.

3 Comments on (new) Simulator Section (3.8) and replies

prepared by Robert Lucas

Comment S005 by JoeSchwarz

p. 1
 3.8.2 Sensitivity calculation
 8.2-R1 It's not clear to me that this is really a simulation; isn't it more an estimate, or an approximate calculation, of the sensitivity levels and dynamics ranges. I don't see any data being *simulated* here. I'm also puzzled by the 0.1-5 minute execution timescale. What kind of calculation could require 5 minutes, even on an ordinary laptop?
 I think that this section ought to be aligned with what's already in 3.3.1-R11. Maybe this section and 3.3.1-R11 could be combined...
 In 3.3.1-R11, there is a requirement there to estimate dirty beams, which doesn't appear here.
 Finally, since the PI would normally specify SNR and get integration time as a result, I think that this calculation needs to be able to go both ways.

Reply:

The sensitivity estimate is not that straightforward as system noise depends on elevation and atmospheric transparency, thus on the hour angle when the observing is done. Also the noise is frequency dependednt across the bandpass ... one might be tempted to do too much here hence the time limit (may be should it be lower like 1 minute?).
 We'll add the dirty beams.
 3.3.1-R11 has been grought down to a mere reference to the present section (as discussed in Granada).
 Doing it both ways is a good suggestion that we accept (get the integration time from the sensitivity required).

Comment S006 by PrebenGrosbol

p.1 3.8.1 8.1-R2

'Relevant parts ... should be available early in the software production cycle ...' . Although this is correct, it is rather a software scheduling issue than a SSR.

Reply:

We still believe it's useful here: otherwise why should we have set priorities in general?

Comment S007 by PrebenGrosbol

p.1 3.8.1 8.1-R3

'schedule preparation phase' suggests a certain operations model where somebody runs and checks simulations before scheduling is done. This seems to be a very complex operations model. If this is required, it should be added to the Scheduler requirements in e.g. 3.4 Dynamic scheduling.
'preferably integrated seamlessly' seems to be a design or/and implementation issue which in principle could be repeated for all parts of the system.

Reply:

- By 'schedule preparation phase' we think rather of long/mid-term scheduling, i.e. as an aid to the staff people that will decide on the schedule of the reconfiguration operations (antenna moves). This long/mid-term scheduling is considered as an input the short term dynamic scheduling, which is the subject of 3.4.
- Seamless integration is a general requirement that particularly needs to be mentioned here (as we have here two rather different software systems that have to be used at the same time in a sort of iterative, interactive way.

Comment S008 by PrebenGrosbol

p.1 3.8.1. 8.2-R1

'schedule preparation' see comment above on 8.1-R3

Reply:

See reply above (007)

Comment S009 by WimBrouw

p. 1
3.8.1

Capability for planning OF OBSERVING PROPOSALS (withreduction), AND TESTING OF DATA REDUCTION PROCEDURES AND THEIR RELIABILITY. Various ...

Reply:

Accepted.

Comment S010 by WimBrouw

p. 1
8.1 -R3
A primary ...simulator DURING OPERATIONS is ...

Reply:

I do not understand, this is not during observations?

Comment S011 by WimBrouw

p. 1
8.2-R1
...should CONTAIN a validity ...

Reply:

I do not understand ??
[validity was in a validation requirement that we dropped in Granada, and thus was no more in the version to be reviewed]

Comment S012 by WimBrouw

p. 1
8.3-R1
- it says 'sensitivity levels'. Is a table expected for say point sources; and for various surface sensitivities?
- it states a variety of parameters; but misses on spectral resolution (or continuum bandwidth; calibration strategies. Is this all embedded in 'configuration? But where is than the spatial resolution taken care of?

Reply:

[8.2-R1]
The plural is just for different spectral resolutions observed simultaneously. We forgot spectral resolution and bandwidth here (to be added). Configuration should be 'array configuration', which is in fact determined by spatial resolution.

Comment S013 by JoeSchwarz

p. 2
8.3-R8 and R9
In both of these sections, we need a bit more clarity. What is the purpose of each simulation (e.g., what parameters of his/her observing program would an observer change as a result of the simulation)? What goes in? What comes out?
In the same vein, array configurations would be *selected* rather than

"simulated". (Although it would be useful for the proposer to be able to get an answer to the question "What fraction of the year will ALMA be in configurations that can be used by my program? I didn't think that was what you had in mind for this requirement.)

The "interactive response" required by 8.3-R8 and the availability to the general user at proposal preparation (8.3-R9) should apply, I think, not (just) to the simulated data, but to the results of processing that data through the pipeline. The proposer, after all, will want to see what kind of image he might get under the hypotheses of the simulation. Do you expect pipeline processing to be available to the astronomer who is preparing an observing proposal with the OT? If so, this has major implications for the software and the need for connectivity to a pipeline service (I don't think that we want to force proposers to load AIPS++ onto their laptops--especially those on which it won't even be supported). Or did you have in mind some kind of "light" pipeline (in which case it needs to be written!).

Reply:

The main parameters the user is expected to change are sensitivity requirement and angular resolution. First time users have a hard time understanding that these two parameters are so intimately related. The simulator is there to help them see the limits of what can be done on a realistic source model.

Array configurations would not be directly selected, but rather selected by the OT on the basis of a resolution requirement from the user. The Simulator should use the configuration as an input from the OT. It's observations in selected array configurations that are simulated, yes.

We expect the simulator to be closely related to the data reduction package (e.g. aips++) in which the data reduction of the simulated data will be done to produced images. This is a subset of the pipeline of course (mostly single-channel data). It seems more realistic that the OT in order to get the simulation done connects to a server that will do the simulation and associated data reduction (that server may be the user's laptop of course). But isn't this implementation?

Comment S014 by PrebenGrosbol

p.2 3.8.3 8.3-R11

'preferably integrated with the off-line data reduction package'
Although it sounds nice, it is not clear if this is the better approach. A separation between the simulator and data reduction package would make a possible to test one with the other and thereby find errors which otherwise would be difficult to locate.

Reply:

I agree with you. We should rather say that the user interfaces are integrated (to hide the actual interface from the user). The actual modules should be well separated (as they are more or less in the WBS).

Comment S015 by TonyWillis

p.2 8.3-R4

The requirement that all Baseline correlator modes should be supported is only given priority 3. While it would probably not be necessary to simulate the behaviour of the entire correlator for proposal preparation (where we probably mostly need information re UV coverage, noise level etc) it will be necessary to do so in order to test the throughput of the pipeline and of the offline reduction package. So I would give this requirement a priority of 1.

Reply:

Following this we would get only priority 1. We agree that the feature is needed, but we need priority ranking. The data reduction developers might want to have it available early for their own purpose, but the ALMA user will most likely not need this at such a high priority.

Comment S016 by WimBrouw

p. 2
8.4-R1

- is the word 'simple' here correct? Simple is ok for the observing tool phase to get to estimates of sensitivities and dynamic range. Simple is not sufficient for data simulation
- .. given observing parameters and models of the instrument, atmosphere AND SKY, INCLUDING SPECTRAL DISTRIBUTION

Reply:

[8.3-R1] - what we list here is not *that* simple, if you add all priority 3 items. Data reduction may want to do more, but this may be a waste of time as we do not know yet many of the *complicated* effects that will appear when the instrument is built and tested.
- add reference to source model.

Comment S017 by WimBrouw

p. 2
8.4-R7

Too restrictive. What about a spectral line profile. It would be essential to be able to model the full line too see the effects in possible broad wings.

Reply:

[8.3-R3] The atmosphere model will compute the atmospheric line wings.

Comment S018 by WimBrouw

p. 2
 8.4-R9
 ... minutes FOR a ...

Reply:

OK

Comment S019 by WimBrouw

p. 2
 8.4-R13.
 - I suppose this too provide a real-life noise and systematic error basis for simple source modelling; and as a base for Monte Carlo experiments?

Reply:

[=8.3-R12] to simulate other, similar data taking and get an idea of the imaging fidelity, but also to simulate how specific observations could be improved by adding data in different configurations.

Comment S020 by PrebenGrosbol

p.3 3.8.3 8.3-R12
 'incorporating' should be clarified. Either is it almost identical to 8.3-R10 or done by the data reduction package which will combine real and simulated data.

Reply:

see 019. We'll explicite.

Comment S021 by RemoTilanus

Aloha Gianni,
 I have essentially no comments on the section '3.8 Simulation'. Goals as for the rest of Alma are laudable but challenging. Given the sophistication of the simulation package proposed, better make very sure it is labeled as 'simulated' data or otherwise recognizable as such.

Reply:

Unfortunately, your comment prevents ALMA from saving a lot of money on hardware ...