February 2015

ASAC Report to the Board

A. D. Bolatto (chair), J. Carpenter, S. Casassus, H. van Langevelde, J. Martín-Pintado, M. Momose, R. Moreno, K. Motohara, R. Neri, N. Ohashi, R. Osten, R. Plambeck, E. Schinnerer, & D. Scott

ASAC Chair rotation

The ASAC Chair rotates every year among NA, EA, and EU. Starting on April 2015 and pending confirmation by the Board the ASAC chair will be the new vice-chair for EA, Dr. Nagayoshi Ohashi. The vice-chairs for EU and NA will be Dr. Eva Schinnerer and Dr. John Carpenter, respectively.

Health of the Observatory

ALMA continues to be successful and productive. The EOC long baseline campaign has been a resounding success, and its SV data has been recently made public. The scientific potential of the long baselines is unique, as already demonstrated by the SV data. We are looking forward to their use in Cycle 3, the first cycle where such capabilities will be offered. Overall, PIs who have received data are happy with its quality, and science observations have become considerably more efficient than they were at the start of Cycle 2, a trend that we hope continues. Nonetheless, many PIs are still waiting for data, and most do not know much about what is going on with their approved observations.
The observatory has stated the goal of concluding "early science" and starting "normal" operations in cycle 5. We applaud the decision of adopting a firm date for the transition.

Follow up of Actions and Recommendations from Previous Meetings

- Urgent action is necessary to implement a PI "duplication checking" mechanism for cycle 3
- The observatory should commit itself to implementing "partial data delivery" for "stale projects" and ALMA large programs (LPs)
- The observatory needs to make public the change request procedure and policy documents
- The data management model should be examined from the standpoint of scientific productivity

Action is needed on making public the change request procedure and the associated policy documentation. This is a very simple action recommended on October 2014 to improve the transparency of the process, and there is no reason for delaying it.

Action is also needed in implementing the partial data delivery plan for the 'stale data' case defined in the October 2014 meeting. **Delays in delivering these data are embarrassing and damaging to the observatory**: the 'poster child' is the largest

approved ALMA proposal, where a significant fraction of the observations were acquired several months ago and the rest is awaiting ALMA to return to the appropriate configuration. During this f2f meeting the JAO and ASAC agreed that these deliveries could take place soon in their simplest form. **The baseline plan is** that raw partial data that can in principle run through the publicly available pipeline should be delivered to the PI when the 'stale data' conditions are met: this requires no extra manpower from the project. The matter of when to start the proprietary period clock for a partially delivered dataset should not become a stumbling block for implementing the partial data deliveries: the "no effort" path is not to start the proprietary clock until all data are officially delivered, so that is our recommendation. If, however, the observatory decides to start running the clock at the time of the partial delivery, we also consider that as acceptable: what is important is that data taken months ago are delivered to the PIs.

We recommend that the scientific leadership of the observatory, with the goal of maximizing scientific productivity, evaluate the end-to-end data management model in use by ALMA, including what is stored in the archive and delivered to the science user (see Charge 3). For example, it makes no sense scientifically that there would be barriers precluding running a partial dataset through the pipeline. **We note that efficient delivery of partial datasets will continue to be necessary throughout the lifetime of the observatory**. There is no "fire proof" scheduling that will prevent the stale data condition from happening in the future (all it takes is a bad storm or technical problems). Also, LPs will require initial "partial deliveries" to fine tune their observing strategy and data products.

Urgent action is needed to put in place a PI "duplication checking" scheme for cycle 3 proposals. The spreadsheet planned by the observatory is a viable solution for cycle 3, as long as all information necessary for duplication checking is available there. We reiterate: **the guiding principle is that no proposal can be down-ranked because of duplications that the PI could not have possibly checked for at the time of submission**. If metadata needed to check for duplications against priority B/C projects cannot be made publicly available in the planned spreadsheet (because they have not been observed or the data are not yet in the archive), **we recommend that no duplication check is run during proposal evaluation against such projects**. The observatory should not alienate the users by asking them to invest significant time in writing a competitive proposal, which then is thrown out because of a duplication they could not have possibly known about. The final solution to this issue lies in implementing a real duplication checking tool, where the planned observation can be checked by the PI against an observatory database without making metadata public. The ASAC alerted the observatory in February 2013 that this was absolutely necessary to fulfill the goal of "no duplicate observations": we are disappointed that two years down the road such a tool is still not implemented.

Charge 1: The ALMA project has emphasized that the last chance to implement capabilities, observing modes and new software for Cycle 4 will be in the first half of 2015. The ALMA project will provide ASAC with input regarding capabilities, new observing modes and software enhancements planned for Cycle 4. ASAC should

review and comment on these modes and give priorities based on expected scientific impact. Are the new modes being discussed sufficient to address the highest impact science themes? Are there other capabilities that should also be given high scientific priority?

Executive Summary

- We agree with the general priorities and timeline for the JAO long-term plan for rolling out new capabilities
- We applaud the proposal of adopting a regular, published, schedule of configurations, and of alternating between "high resolution" and "high frequency" years
- We recommend that top priority is granted to capabilities that have the potential to increase the observing efficiency of ALMA
- We recommend that the lowest priority among the "medium risk" modes identified by the JAO is assigned to TP B9/B10 and TP continuum observations
- There is urgent need to put together an ALMA mm-VLBI policy. This is a fundamental step to offer such a capability, which could be done in cycle 4. We stand ready to provide comments on the plan on a short timescale

The observatory plans to enter a scheme where a "fixed" configuration schedule is adopted for the first 5 months of a cycle (fixed at the call), with the remaining 7 months loosely determined at the time of the call and fixed after the proposal review. This will improve the planning and predictability of the observing, and make the result of the proposal review more realistic in terms of "executability" (more on this later). By advertising it to the users before proposals are due it will allow them to optimize the proposed sources, and by making it available to the reviewers it will improve judging the feasibility of proposals. Because June-August are the best months for high frequency, and June-November are the months where the phase stability allows observations in the longest baselines, the observatory proposes to alternate between years of compact configurations in June-August (best for high frequency) and years of extended configurations. There will always be extended configurations in September-November. We think this will help optimize the science output while permitting the observers to plan ahead, so **we applaud this decision**.

Deputy Director Stuartt Corder presented the ASAC with a detailed plan for the roll-out of capabilities that breaks them into "low", "medium", and "high" risk according to their difficulty and the possibility of causing disruptions while implementing them. The planned capabilities are as follows:

- Low risk: efficient spectral scanning, spectral polarization (e.g., Zeeman effect), OTFi, ACA standalone proposals, multiple mosaics, multi-intent sources, calibration sharing across observing blocks in pipeline, daytime B7

- Medium risk: solar observations, 90-deg switching (for sideband separation in B9/B10), single dish continuum, single dish B9/B10, VLBI, ACA linearity improvements (improved TP calibration)
- High risk: expanded polarization, higher spectral resolution, linearity in baseline correlator, ACA correlator subarrays

The longer-term plan (cycles 5 and 6) includes multi-resolution correlator modes, improved polarization, band 5, dynamic subarrays, and frequency switching for TP among other capabilities.

The plan presented by the JAO is detailed and well-thought out. We maintain the advise of focusing first and foremost on improvements that will increase the efficiency of the observatory. At the request of the Deputy Director to identify the highest and lowest priority items, we identify single-dish B9/B10 and single-dish TP as the lowest priority "medium risk" developments for cycle 4, one of which may be delayed to cycle 5 if problems arise. A criterion for choosing among these two is to evaluate the potential user pool by looking at the number of proposals requesting ACA for B9/B10 and proposals requesting continuum for the lower bands in cycle 2/cycle 3. We also identify "90 degree switching" as the highest priority in the "medium risk" category.

The longer-term plan contains two improvements that bring unique science capabilities and may be raised in priority to be offered earlier. Science subarrays will allow the use of subarrays to more accurately measure the SED of rapidly variable sources (e.g., Solar flares). Precision polarimetric calibration will enable difficult observations that only ALMA can do (e.g., Goldreich-Kylafis effect, polarization mapping over large areas). We also note that mixed-resolution correlator modes may allow some gains in efficiency for specialized observations.

If the commissioning proceeds according to the plan, **mm-VLBI should be offered as a capability in cycle 4**. Letting it slip to cycle 5 **would be unacceptable**. In order for mm-VLBI to be offered, **a policy needs to be worked out as soon as possible**. We are pleased to hear that the observatory is working at putting together a planning document, following the Bologna meeting. Robert Laing presented the ASAC with the highlights of the document, which is still in its early stages. We strongly support the principles we were presented with. In particular, the idea that the VLBI proposals should be evaluated together and competitively with "normal" proposals, with no especial time slice set aside for VLBI echoes the thoughts of the ASAC. Also, we very strongly support the archiving of calibrated visibilities and images in the ALMA archive for all participating telescopes. We note that productivity (in terms of publications) need be an important consideration when awarding ALMA time, and VLBI should be no exception. Given the possibility of offering mm-VLBI in cycle 4, **there is urgency in completing the planning document**. The ASAC should be part of the process; **we are looking forward to commenting on a complete version of the VLBI plan** on a short timescale. We also note that in order to be able to offer this capability in cycle 4 the OT has to be ready to accept this type of proposal.

Charge 2: Pursuant to standing Charge 2, continue to assess the status of Cycle 1 and Cycle 2 observations. Are there improvements in execution efficiency and does the projected program completion rate coincide with the expectations of the user community? Are the data being released to the PIs in a timely fashion, and are adequate progress updates being communicated to the PIs and the community at large? ASAC should continue their commentary on ways to improve the data release process.

Executive Summary

- The observing efficiency has improved greatly with respect to what it was at the start of cycle 2
- The observatory has made a serious effort at evaluating the completion fraction of cycle 1/cycle 2 projects
- Because of scheduling inefficiencies and limited time, the likely completion fraction by the end of cycle 2 is ~83%, **if there are no major weather or technical events**
- We strongly recommend that to avoid the risk of falling below 80% action to increase the amount of science time be taken now. This can be reduced close to the end of the cycle if no major problems occur
- We suggest contacting PIs to see if the scheduling efficiency can be improved
- We strongly advise that to avoid similar situations in future cycles it is imperative to connect better the acceptance of proposals and their projected executability
- We are concerned about the important, and growing, backlog of unprocessed data despite the use of the pipeline

The ASAC received the JAO report by Lars Nyman. It is very comforting to see that the JAO efforts at improving the system stability and the handling of antenna integrations have paid off, resulting in a much higher efficiency (60% vs. 36%) and higher number of antennas available: we congratulate the observatory for the success, and for making these improvements a priority.

We were presented with a detailed evaluation of the fraction of projects that will be likely completed as a function of configuration. The mismatches inherent to the fact that some sources do not transit at night during the configuration requested for their observation add up to important scheduling inefficiencies (Figure 1). The bottom line is that, assuming the observatory can maintain the high mean efficiency of the last several science observing blocks, the fraction of projects that are likely to be completed is estimated at ~83%. **The ASAC is concerned that there is no contingency in this number**: a bad snowstorm can result in ~70% completion efficiency, and a bad winter can certainly drive that number even lower.

Although ALMA still operates under a "best-efforts" policy, **it is key for the credibility of the observatory among the science community that the final**

completion fraction of Cycle 1/2 projects is not significantly below 80%. To have a reasonable chance at reaching this goal, we strongly recommend action be taken now to maximize the fraction of science time. This can be reversed later if the progress at completing projects is good, while the converse may not be possible. Since sub-arrays will potentially enable some engineering tasks to be done in parallel with science observing, getting them operational is to be encouraged. We endorse the JAO plan of contacting PIs of projects that are unlikely to be completed to warn them of that fact in time for them to resubmit a proposal for cycle 3. Moreover, we suggest that the observatory considers contacting PIs to see if the configuration requirement can be in some cases relaxed without loss of science, to allow for a better match of sources and configurations, reducing the scheduling inefficiency, and improving the completion fraction.

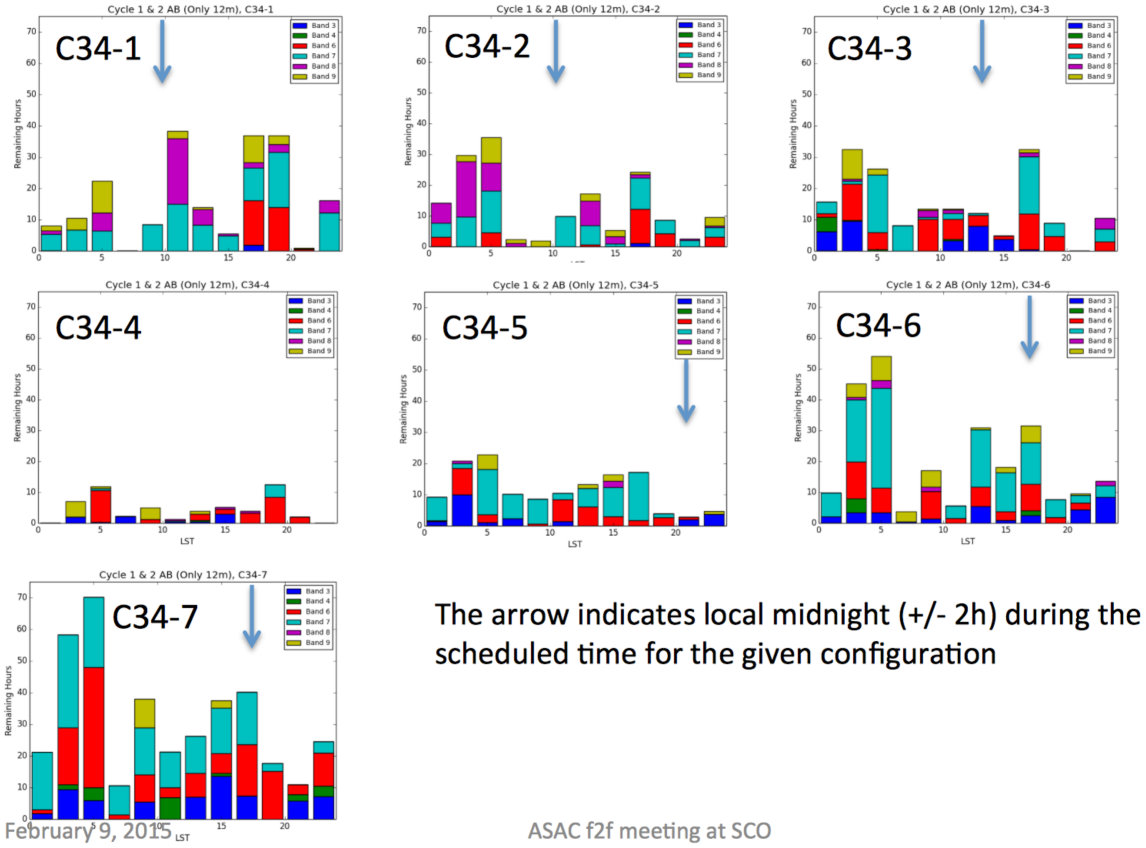

Figure 1: The plot shows the hours that still need observations in different configurations compared to the midnight transit time. There is a fundamental scheduling inefficiency that needs to be avoided in the future by matching accepted projects to the schedule of configurations (determined by weather).

It is crucial for ALMA to avoid falling into a similar situation in the future. Besides the technical difficulties experienced, this situation is fundamentally caused by a poor match between projects accepted purely on scientific merit and the reality of weather limitations for mm-wave observations (Figure 1). This problem will only become more prominent with longer baselines. The reality is that the observatory

can only afford to accept very few "hard to execute" projects, otherwise risk disappointing the users.

The idea of having a "semi determined" schedule of configurations will help, but the observatory needs to evaluate how many hours at what LSTs and frequencies can be possibly accepted. ASAC agrees with the observatory that introducing the division of standard/non-standard modes, as well as disclosing some matrix of execution rate statistics, is useful in order to make proposers more pragmatic. But we are concerned that this may not be sufficient to avoid a similar problem in Cycle 3 and beyond. **We recommend the observatory is proactive at inserting some measure of "executability" in the process of selecting proposals**. A possible mechanism is to run a mock schedule including realistic weather statistics before proposals are accepted, but simpler options limiting the number of hours in particular RA and weather ranges would also work.

The ASAC is very concerned about the important (and growing) backlog of unprocessed data, as discussed by Lars Nyman. We note that there does not seem to be a measurable improvement in data delivery rate even after the introduction of the pipeline, and although the EOC Long Baseline paused allowed the JAO to almost catch up with the science data backlog, the gap is rapidly widening again. The observatory is also concerned about the backlog, and has already prepared a plan to mitigate this situation, which was discussed by Eric Villard. In particular, stricter/more detailed guidelines in the use of the pipeline may result in efficiency gains. It is key to monitor the data delivery over the next few months in order to establish whether there are real improvements. If this does not appear to be the case, then further actions will need to be considered.

Charge 3: From their own experience and from community input ASAC should investigate and comment on the "face of ALMA" to the scientific community.
1) Is the information available through the ALMA Science Portal sufficient to enable users to complete observing proposals and successfully carry out the science goals if their program(s) are accepted? Is the Portal sufficiently informative for astronomers who are not experts in (sub)millimeter interferometry? Are there ways to improve the clarity and/or content of the Science Portal?
2) ASAC should comment on the ease of using the Project Tracker to follow progress of approved programs and suggest improvements to JAO where necessary.
3) Does the archive query tool offer the correct capabilities for PI access to proprietary data and for general user access to archival data? Is the archive sufficiently user-friendly for scientists who are not experts in (sub)millimeter interferometry?
4) Are there other areas of interface with the user community that could be improved?

Executive Summary

- The archive is a key part of the observatory. It is crucial that it is of use for data mining and friendly to non-expert users
- We are concerned that key data, such as approximate angular resolution and sensitivity and whether it is a 12m or 7m dataset, are currently missing from the information provided by queries
- The project tracker interface needs to be simplified and designed to be easy to use and understandable for PIs. We will make specific recommendations to the project
- The automatic email system needs to be made much more informative
- We would like to see a plan for "future usability improvements" of the OT provided by the observatory
- The Science Portal is reasonably user-friendly and well organized
- We recommend the next ALMA conference is organized by NA sometime near middle of 2016

The ALMA archive is a key part of the observatory, not just from the standpoint of storing the PI data. Analysis of productivity of mature facilities shows that publications from archival data can rapidly overtake original PI observations. Thus the archive may ultimately determine the scientific productivity of ALMA. With that in mind, it is crucial that both expert and non-expert users can mine the archive for data.

Felix Stoehr presented the ASAC with an archive development plan. Some of the recent accomplishments include programmatic (non-interactive) query access through the astroquery library in python. A medium-term goal of the development is to provide users with a complete science-product archive including products and previews for every source and spectral window. Given the incomplete processing of early science data, the archive will focus on the short-term development of previews for early ALMA observations. **The ASAC agrees that making science products is an excellent goal that deserves support**.

An important concern for the ASAC is that the observation information currently resulting from queries is incomplete. For example, there is no information about the range of spatial frequencies or the approximate angular resolution of a dataset. Surprisingly, there is also no information about what type of data is contained in a set (12m, 7m). Finally, there is no sensitivity figure (there is an integration time that is often in error). This is information that is necessary to establish whether an observation is a duplicate, and it is key for scientists mining the archive or deciding whether to propose for a source.

The structure of the data retrieved from the archive is complex, with many subdirectories on several levels of depth, and obscure naming conventions (e.g., science_goal.uid___A002_X5a9a13_X711). With any program of moderate complexity, **this obscurity makes it unnecessarily hard for the user to keep track of precisely what is what**: it seems that an intuitive naming convention

(perhaps based on the names of science goals and sources) could be used and it would be strongly preferred. A flatter directory structure that generates many levels **only when necessary** would also be much more user-friendly.

The archive currently stores raw data, scripts, and calibration tables. The ASAC did not discuss the long-term viability of the archive this time around, but **have expressed concerns before about the fact that calibration scripts will only run on a particular version of CASA**. Thus, maintaining a viable archive over the lifetime of ALMA (and longer) implies an indefinite commitment to providing users with the possibility of executing scripts in old versions of CASA.

The Project Tracker is a tool that was developed for internal use of the project, which has been adapted as a PI communication tool. Unfortunately, the PT interface has a wealth of confusing information **that makes it extremely hard to use for the non-initiated**. The ASAC has recommended a simplification of the PT interface. At the request of the project a subcommittee of the ASAC will suggest specific improvements.

A consistent problem is the lack of information regarding progress on projects experienced by most PIs. PIs can currently sign up for the (optional) email notification system, which triggers automatic emails at different stages of the process (currently when a dataset is completely observed, or when it passes QA2). An outstanding problem is that these emails are extremely non informative. A verbatim example is:

```
The project components for ALMA Project  have undergone the following
state changes:  OUS "Member OUS (NGC253)" state changed to
FullyObserved.  You can access the Project Tracker under the Observing
menu item in the Science Portal to follow up the progress of your
observations: https://almascience.org/observing/project-tracker  Kind
regards,  The ALMA State System
```

With this information is impossible to understand precisely what has been observed. To be an **effective communication tool**, these emails need to be phrased in clear terms, sent every time observations are acquired by default, (so the PI feels that things are moving) and communicate precisely what setup has been run. The PI should not need to proactively select to receive email, but could select not to receive it. Coupled with a published configuration plan, this would be an excellent way of letting PIs know what is being done when.

Andy Biggs presented the ASAC with a short summary of improvements to the OT for cycle 3. These are a mix of functional (e.g., better source query) and usability (e.g., import/export of pointing positions) improvements. The ASAC would like to see a **long-term plan for future usability improvements prepared and provided by the observatory**, to be able to evaluate whether they are sufficient or need to be tweaked.

We had no important comments on the Science Portal, **which seems fairly well organized and easy to use**. A minor comment is that the user authentication seems to expire fairly quickly (on the scale of hours), which requires the user to log in several times per day. Making the user log in only once and maintaining the authentication for several days is unlikely to cause security problems.

The ALMA conferences are a very important part of the "face of ALMA" to the external community. The ASAC recommend that **the next conference is organized in NA** (maintaining the rotation between the regions), and that it takes place sometime around middle of 2016, which should allow PI long-baseline results to be presented.

Charge 4: ALMA Development Plan Standing Charge 4: ASAC should complete their work on the ALMA2030 document, submit it to the Board and to JAO, and present a report at the March 2015 ALMA Board meeting.

Executive Summary

- The ASAC recommends a number of mid and long-term development paths for the ALMA observatory focusing on better archive and tools, improving Rx and wider bandwidths, longer baselines, and long-term research on wide-field imaging

The ASAC converged on a set of big "development themes". The result is a mix of mid-term and long timescale development effort, some of which overlap with studies currently underway. The three documents that form the foundation of the ALMA 2030 report (Pathways for Developing ALMA, Major Facilities, Science Themes) will be complemented by a fourth document discussing the conclusions of the study. **The entire study will be published as part of the ALMA memo series.**

The ASAC notes that, separately from the development paths recommended below, a large single-dish telescope equipped with cameras capable of fast large-scale mapping would be an important scientific complement to the interferometer. Such an instrument is outside the scope of the envisioned ALMA development projects, but if built there would be large potential scientific synergies with ALMA (for example, surveying for sources, or providing the larger source context).

The consensus recommendation, after the discussion that took place at the f2f, falls into four categories:

1. **Better archive, mining, and analysis tools, enabling gains in usability and impact for the observatory**. The archive should ideally contain full science-ready products and value-added products resulting from automatic data analysis. It should be VO compliant, enable efficient mining, and allow for user-submitted products to enhance the impact of the observatory.

2. **Wider bandwidths and improved receivers, enabling gains in speed**. A long-term goal should be to be able to correlate an entire band in one observation. This requires improvements to the receiver IF, the IF transport, the correlator, and higher data rates. Simultaneously, it is key to continue work on improving receiver sensitivities.
3. **Longer baselines, enabling qualitatively new science**. A goal is to progressively increase the longest baselines, starting with doubling the current maximum baseline. It is important to evaluate how far real-time correlation can be practically used (instead of VLBI techniques such as fringe-fitting). An important aspect of this improvement is continued work on atmospheric phase correction schemes, improving coherence on the longest baselines.
4. **Wide-field imaging, enabling efficient mapping**. It is technically challenging, although probably feasible in the long-term, and it needs up-front research. It requires in-depth investigation of multi-pixel receivers for interferometry and their implementation in ALMA. It is likely practical for only one band, optimally a high demand band with limited FOV such as B7 or B6. It requires wider IF transport, correlator, and data rates as with "wider bandwidths" above.