CASA Users Committee Report 2014

October 23, 2014

Executive Summary

In this report, we present comments and suggestions resulting from the inaugural face-to-face meeting of the CASA Users Committee (CUC). The CASA team continues to do an excellent job providing software to process data from ALMA and the EVLA, essential to realize their science potential, especially as the capabilities and requirements rapidly evolve. The CUC applauds these efforts, and notes that the CASA team does a very good job with the challenging task of balancing telescope commissioning needs with user support. Here we briefly summarize some of the key feedback on our charges on the pages that follow:

- Since speed and ease-of-use are the main laments of CASA users, we recommend to emphasize improvements in performance and parallelization, and to improve the overall robustness of basic operation. Additional attention to setting user expectations may also help ameliorate these issues, such as (crude) time estimators for critical tasks.
- Complete the CLEAN refactor with high priority, as this will bring important capabilities as well as retire some speed and usability issues.
- Improve and curate better CASA documentation; a small number of well-maintained documents is preferable to many that are not maintained and become out of date.
- Provide updates to CASA users about ongoing and planned developments, as well as prioritized fixes, in a regular and easily digestible form.
- Improve feedback to users on helpdesk/ticket disposition, especially where challenging problems may take a long time to address, or will not be addressed.
- User surveys can uncover problematic issues and help to guide future development; we would like to work with the CASA team to develop and deploy more user surveys.
- As there always will be a longer wish list of CASA improvements than available resources can support, we recommend to engage the external community by stabilizing and documenting interfaces and import/export formats to better enable users to contribute and to maintain CASA tasks and scripts.
- Continue to promote CASA for telescopes other than ALMA and the VLA on a best effort basis; major initiatives formulated by external groups should be undertaken only if the additional resources required are provided.

Introduction

The CASA Users Committee (CUC) was formed to provide independent feedback concerning the perceptions of the entire user community of the Common Astronomy Software Applications (CASA) software package on capabilities, usability, reliability and performance, with the primary goal of offering insights into user needs to maximize science from ALMA and EVLA. The CUC also offers strategic input on large-scale or long-term initiatives of CASA development, related computing infrastructure, and support for telescopes other than ALMA and EVLA. The CUC is tasked to interact with the broad community of CASA users. This interaction is meant to formulate recommendations that supplement, not replace, the official communication channels between CASA users and developers, such as the help desk. The CUC respects CASA management priorities and does not act as an executive body.

This report contains responses to the charges presented to the CUC by Brian Glendenning, the head of data management and software development at NRAO, for the first face-to-face meeting of the CUC with the CASA team. This meeting was held September 22–23 at the Pete V. Domenici Science Operations Center in Socorro, NM. The membership of the CUC consists of representatives from the ALMA partners. The full committee was in attendance for the meeting, consisting of:

David Wilner (NA, chair)
Rachel Akeson (NA)
John Carpenter (NA)
Thibault Cavalié (EU)
Rachel Friesen (NA)
Lizette Guzman (CL)
Tomoya Hirota (EA)
Daniel Jacobs (NA)
Alexander Karim (EU)
Kazushi Sakamoto (EA)

At this meeting, the CUC was presented with extensive information from the CASA team on a wide range of topics concerning ongoing developments, challenges, and plans for the future. We express our gratitude to all of the staff involved in this endeavor for their considerable efforts. Our resulting recommendations and comments are based on the reflections and expertise of the individual CUC members, on input from individual CASA users (mainly ALMA and EVLA users), the ALMA regional centers, as well as the ALMA Cycle 2 user satisfaction survey provided to us, and a pilot poll of Facebook astronomers (initiated by CUC member D. Jacobs).

1 Capabilities, Usability, Reliability, Performance

Please comment on the capabilities, usability, reliability, and performance of CASA for current users of ALMA and the EVLA. Should any of these areas be getting a significantly different allocation of effort than is currently the case? What areas would you de-emphasise as a consequence?

The CUC received extensive reports on plans to introduce new capabilities into CASA and on ongoing steps to improve the performance of CASA. Much less emphasis was placed on reporting on usability, and very little information was provided about CASA reliability.

1.1 Capabilities

The capabilities to support basic calibration and imaging of ALMA and EVLA data are essential and must have highest priority. This is proceeding well. For example, the CLEAN refactor underway will provide needed capabilities for upcoming observations, as well as simplify user interaction with a heavily used task. For now, we recommend that relatively little emphasis be put on capabilities that are readily duplicated in other software packages, such as the production of publication quality figures. For this strategy to be most effective, CASA data import/export tasks must be clearly documented.

1.2 Performance

The CASA team is devoting considerable resources to improve CASA performance through parallelization and optimizing disk reads, and the CUC was impressed with the results of this ongoing effort. The CUC agrees with the high priority placed on these developments, as the time it takes to reduce data with CASA is one of the primary complaints from users (see §4 for additional discussion of ways to ameliorate the widespread perceptions of CASA slowness). A related area of concern to many users is the computing hardware required to run CASA on their ALMA and EVLA datasets, and we would like to ensure that the on-line hardware recommendations are updated as appropriate. This will require close monitoring going forward given the rapid evolution of the size of ALMA/EVLA datasets, CASA, and the hardware available.

1.3 Reliability

We recommend that steps to improve CASA reliability should receive more emphasis. The experience of CUC members is that CASA throws frequent segmentation faults or other errors that can be extremely frustrating. While restarting CASA is often sufficient to proceed, such errors undermine general confidence in the entire software package. The overall robustness of CASA is deserving of careful attention and can be improved.

Another area that is no less important than preventing segmentation faults is to avoid implementation errors. These may not crash CASA but silently result in erroneous outputs, or outputs different from a user's expectation. For CASA serving both ALMA and EVLA, it is essential to have sufficiently wide range of usage examples and situations for both

telescopes for design checks and testing. Since individual programs/tasks are often written by a small number of people, or a single person, there is natural concern that the variety of usages considered may be restrictive. We therefore suggest to take (or continue) measures to reduce the chance of oversight due to the small number of eyes on each CASA program/task.

We recognize that the presence of some bugs is inevitable. For those that are known or found in the released versions of CASA, it is important to inform users about them and possible workarounds. This helps CASA users to have a sense that their reduction results are reliable. In this regard, we welcome that the CASA team has been maintaining a list of known issues (http://casa.nrao.edu/release_ki.shtml). We make some suggestions in §4 to further improve the information flow. We note that information about past bugs is also important, in order for users to judge the reliability of previous data reductions made with an old CASA version, and when users may need to revert to an old CASA to reduce archival data.

1.4 Usability

We recommend that more attention be placed on the maintenance of the documentation that users rely on. While the on-line CASA guides are generally up-to-date, knowledgebase articles often refer to old versions of CASA, and it is unclear if a problem (and solution) noted in a knowledgebase article is appropriate for the recent release. The CASA help page returned by Google includes links to a Release Notes for CASA version 3.0.2, documentation for the ALMA simulator in CASA version 3.4, and other outdated links. The online documentation of the CASA toolkit also appears to be incomplete and out of date, at least in parts. Obsolete pages really need to be clearly marked or removed. Our general recommendation is that it is not necessary for the CASA team to produce *more* documentation, but to improve and *curate better* the existing documentation.

We also suggest that the growing CASA user base would be well served by better communication from the CASA team, with regular updates of ongoing and planned developments, as well as prioritized bug fixes, in an easily digestible form. This improved transparency will aid users in planning their ALMA and EVLA projects, as well as increase overall confidence in the CASA package. One possibility to accomplish this might be a CASA Newsletter, analogous to the AIPS Newsletter, although there are many possible options. (If Web 2.0 features are used, then on-line user comments may prove a fruitful resource to users and developers alike.)

The support for Mac OS platforms appears to require considerable resources from the CASA team. While the CUC was informed that half of CASA downloads are for Mac OS, it remains unclear if the primary usage is for laptops or for desktops, how much Mac OS is relied upon for data reduction, and if the data reduction usage will change in the future as the datasets continue to increase in size. We recommend that Mac OS usage of CASA be monitored closely (perhaps through user surveys), to determine the level of future support required for Mac platforms.

If upcoming ALMA and EVLA datasets generally cannot be reduced on the laptops and typical desktop systems available to users at their home institutions, then more resources will need to be deployed to offer other options for CASA processing. One possibility may be a remote client arrangement for centralized computer clusters; in that case, the user access

and capacity of the cluster infrastructure to meet these needs will need close examination. We wonder if the throughput of ALMA and EVLA science already is being affected by local data processing limitations of users.

While scripting CASA is clearly an increasingly popular and prevalent way of using tasks and tools, there are still users, situations, and tasks that benefit from an interactive front-end. We make no recommendation to change the balance of support for these various user modes at this time.

2 Relative Priorities of Functional Areas

Similarly, would you change the relative priorities of the functional areas: Single dish, imaging, calibration, visualisation/analysis?

Calibration and associated imaging of ALMA and EVLA data must have highest priority. We emphasize that calibration development is naturally tied closely to the release of ALMA and EVLA capabilities, especially as these capabilities are advertised and made available to large numbers of users. High priority should be put on automatic calibration procedures in pipeline processing for the basic capabilities *that are most in demand by users*.

Making large synthesized images with multiple thousands of spectral channels is a challenging and time-consuming task. The CUC praises the CASA team for extremely important work in imaging development, coordinating with the schedule of ALMA and EVLA needs. While the CUC did not discuss specific priorities in imaging capabilities, all of the multi-term, multi-frequency synthesis, multi-scale synthesis, and related developments are clearly important, and these likely would benefit from additional (albeit highly specialized) effort. In particular, the reliability of new imaging processes must be tested extensively.

One area where users express a clear demand for improvement is for clarity in the visibility import/export tasks. Since CASA does not yet support modeling of visibilities beyond simple (Gaussian) features, an important subset of users need to analyze visibilities outside of CASA, using other software or their own analysis tools. Also, some users would like to use CASA to analyze/visualize visibility data produced by non-native systems. Since the visibilities are the basic observational data of ALMA and EVLA, whatever format is used by CASA for import/export (e.g. standard FITS, simple tables, etc.) must be tested and fully documented for this purpose. Ongoing changes in the expression of visibility weights have been a source of confusion, where users have found answers difficult to obtain. We note that the users who write visibility analysis software are often ones who are pushing boundaries and have the ability both to recognize subtle software problems and to provide useful feedback about them.

We heard about CASA development to analyze data from single-dish radio telescopes, in particular driven by the needs of ALMA total power observations. While these single-dish developments are undeniably important, we sensed that a larger relative effort was put toward this than to other, perhaps more pressing, needs.

Given that the ongoing new requirements of ALMA and EVLA data retain top priority, it is wise that the CASA team is both willing and able to leverage outside development projects to provide users with more complex analysis and visualization tools. The CARTA

(Cube Analysis and Rendering Tool for Astronomy) software, realized as a collaboration with University of Alberta as an ALMA Development Project seems to be a good model. CARTA has the potential to provide a next generation drop-in replacement for the current CASA viewer, with scripting capability. This CARTA development also highlights the fact that the CASA team needs to prepare users an upcoming transition in visualization capabilities by providing more (and more regular) information on the remaining plans for the CASA viewer and plotms, the promised capabilities of CARTA (and gaps in its capabilities), and how CARTA will work within CASA.

To better enable external users to develop analysis tools for use on ALMA and EVLA data within CASA, the CUC would like the build process of CASA to be open, which would allow users to add modules and plug-ins. Taking advantage of the user community to help create such tools not only has the potential to benefit a broad spectrum of users but also to allow the CASA team to focus resources where they are most needed to deal with the requirements of ongoing instrumental developments.

3 Pipelines

While the prime motivation for the pipeline development is for institutional use, it will also be delivered to users. Do the pipeline capabilities appear to be appropriate for that use? Would you suggest any change in emphasis of the pipeline vs. CASA as a traditional radio astronomical data reduction package.

The pipelines should be made available to all users. This is especially important for the population of less expert users who benefit most from automation. Getting robust and reliable calibration pipelines for ALMA and EVLA fully tested and properly delivered is a high priority. We recognize that imaging pipelines are more challenging, given the wide variety of science applications and circumstances of individual programs, so we suggest to keep these simple and focus on basic features. Once the user community gains sufficient experience with using these pipelines, then we can evaluate usage patterns to help establish priorities between pipeline and "traditional" CASA support. We believe it is premature to comment on this now.

The pipeline documentation effort is an important component of pipeline development. This should be guided by early user tickets and questions. As already noted, the CUC believes that it is better to have a small number of well-maintained documents than to have many documents that are not maintained and become out of date.

With respect to ALMA, we would like to see the backwards compatibility of the pipeline to past datasets (mainly Cycle 0 and 1) more clearly defined. For processing archival data, the CASA group may need to work with the ALMA ARCs to ensure that the relevant versions of CASA are available to run calibration scripts archived with the data (that cannot be run through the pipeline). The responsibilities for this code maintenance need to be specified among the responsible parties.

4 Barriers to Adoption and Community Input

Are there remaining barriers which prevent CASA from being more widely adopted, or is it now a matter of personal preference? Are there any mechanisms you would suggest (other than the CUC itself) to more systematically solicit the viewpoints of the broader community (e.g., a survey).

While the CUC membership is diverse and represents a variety of interests, our knowledge of typical user stories and use cases is still limited, and we believe these questions will be most properly addressed by user surveys. We would like to work with the CASA team to develop future user surveys.

At the moment, we have access to two recent user surveys that inform us about CASA adoption. First, ALMA asked a large number of questions about CASA as a part of an more comprehensive survey of 4042 Cycle 2 proposers. Second, CUC member D. Jacobs initiated a short survey of the Astronomy Facebook group ($\sim$ 1200 professional astronomers). Response rates were low but respectable, 10% for ALMA and 3% for Facebook. In absolute numbers these two surveys together netted 452 responses. For context, each version of CASA is downloaded about 1500 times. Both surveys showed that about 65% of the respondents use CASA. The two most common items cited as either complaint or barrier to CASA use may be summarized as (1) CASA is slow (20-25%) and (2) CASA doesn't suit my needs. There are also issues with interactions between users and the CASA team.

4.1 Speed

We speculate that there are two main camps of users complaining of slowness: (1) those not acquainted with the large sizes of new interferometric data sets, particularly new ALMA users, and (2) those impacted by current limitations of CASA, such as requirements for disk I/O capabilities. The first group is the most likely to express frustration with extremely large datasets, complex documentation, and the overall "speed" of processing. Novice users may generally prefer a pipeline product, especially if they are unable to process data locally due to lack of experience or limited computing power. The second group may be evaluating CASA against historical expectations and the performance of other packages such as MIRIAD or AIPS, without careful consideration of the algorithms, input parameters, and hardware optimization. Such detailed comparisons are generally not available. We suggest that some of this dissatisfaction might be alleviated if key CASA tasks provided simple estimates of the run time expected based on the input parameters and hardware configuration. Providing users with some benchmarks for common processing tasks also would be useful.

4.2 Needs

Some users would like to be able to access their favorite tasks from other packages at the same or better efficiency in CASA; they see the absence of these tasks or features and resist switching to CASA. Others may have more specialized needs, including non-standard observing modes of ALMA or EVLA, or other telescopes. The CUC's attention was often called to the conflicting interests of developing vanilla pipelines that serve a very large user

base, and providing a robust foundation for the general analysis of interferometric data. It would help serve this latter group if software interfaces are clearly defined and timely information is provided about the evolving landscape of CASA developments.

4.3 User Interaction

The current user-CASA interaction process requires direct one-on-one conversation with the helpdesk for every user, for every problem, regardless of how often the problem is reported. Typical questions that often cannot be easily answered by a user confronting a problem with CASA are (1) Has anyone seen my problem before? (2) Has my problem or request been sent to developers for attention? (3) Where does my request stand on the development timeline? We emphasize that an answer along the lines of "can't fix it" or "feature waiting on resources, 6-12 months horizon", is preferable to the current lack of information flow.

Actual CASA bugs and feature requests are submitted to the internal ticket system (JIRA) for tracking through time, and this ticket system is used extensively by developers and is thought to provide a good snapshot of the state of CASA. While we understand the resistance to making public the internal ticket system, we urge that some form of summary information be made available (such as the titles and dispositions of the tickets). Similarly, we urge access to summary results of the CASA prioritization process that results from the periodic meetings between developers, management, and ALMA and EVLA scientists that fold new requirements, feature requests, known bugs and long term needs into plans for the next CASA development cycle. (Perhaps a CUC member could be included as an observer in this process, who could provide a short report on the outcome in some form of regular update to the users, see §1.4.)

5 Support for other Telescopes

Although funded by NRAO and ALMA to support ALMA and the VLA, CASA/casacore are also used in various capacities by a number of other current and future radio telescopes. Does the CUC have any suggestions about whether and how these efforts could be better supported without negatively affecting the core CASA mission?

CASA offers enormous potential for many telescopes, and not only radio telescopes. For example, good visualization and analysis tools can attract a far larger base than just traditional radio astronomers. The CUC regards the current flexibility of CASA in this regard as highly valuable. Of course, we recognize the limited resources of the CASA group and the priority that must be given to the evolving needs of ALMA and EVLA. The diversion of significant resources to external groups is clearly not in the best interests of the majority of users that the CUC represents. Hence we encourage continuing the policy of CASA support of other telescopes on a best effort basis, as it seems to be working well without drawing too many resources from other priorities. Major changes formulated by external groups should be addressed only if they provide the resources (funding and manpower) needed to implement these changes.

The CUC envisions that other observatories could be enabled better to autonomously

integrate CASA if the interfaces needed by external parties are standardized and fully documented. We suggest as a goal to have a well-defined framework for external groups to utilize, avoiding individually tailored solutions where possible, in order to minimize the workload on the CASA team. In this context, it should be clear in the documentation and directly on the CASA helpdesk front page that the helpdesk is amenable to submissions by users of the wider group of telescopes using CASA, to deal with issues that remain.

6 CUC Management and Organization

The CUC should appoint a deputy chair, and might also choose to suggest other organisational improvements consistent with its terms of reference.

The CUC has elected Thibault Cavalié as deputy chair. The intention is for the deputy chair to back-up the current chair as needed and to ascend to the chair position in advance of the next face-to-face meeting.

If desirable, the next face-to-face meeting to be held earlier in the CASA development cycle, to increase the impact of the recommendations. The CUC plans to continue to hold telecons at roughly quarterly intervals.

The CUC as initially composed is relatively thin on expertise with low- and mid-frequency EVLA observations. We have reached out to colleagues with appropriate experience for their input. However, given the importance of this regime and its unique challenges, for example with interference mitigation and wide-band, wide-field imaging, it is highly desirable to include additional representation in this area as soon as possible.