

## **NA Position Paper on PI “immediate access” to raw data.**

*Submitted by NA ARC Manager on behalf of NRAO & NRAO Users Committee*

Several of us were involved in writing the ALMA Operations Plan, and our intent was for PIs to have immediate access to their raw data (*see accompanying “Extracts from AOPvD”: Sec.4.4 and the paragraph from Sec.4.13 associated with footnote 15*). For Early Science we agreed that it was reasonable to suspend this requirement, and to provide observational data to PIs only after all data have been calibrated, imaged, and quality assessed by observatory data reduction experts to prevent poorly characterized data from being sent to PIs. But the steady state model should remain as originally intended: PI access to the raw data from individual executions as soon as it is available in the regional archive. The delivery of quality assured data products would still follow after all observations were complete, and the proprietary period of all associated data would be based on the delivery date of those products. That is, the availability of the raw data would not count against the proprietary period. Raw access should be provided on an “as is” basis; no data reduction support apart from the already existing CASAguides tutorials should be offered to PIs so as to not unduly burden the support staff at the ARCs. The above model requires very little changes to existing ALMA software and procedures.

In their 2013 report, the NRAO Users Committee expressed their preference to allow ALMA PIs “immediate access” to raw data:

*"To avoid random delays in accessibility to the data and to allow the PI to assess the utility of data flagged as “poor”, the UC suggests that raw data be distributed into the ALMA archive immediately for the PI to access. Pipeline-processed products/QA2 can be delivered as an additional component whenever they become available at a later time. This would also leverage more “experienced” users in the community to identify problems and reduce the stress of the staff to shorten the timescale of QA2."*

This echoes a sentiment we have heard at various science meetings, that the observatory should not prevent access to the raw data for their projects, which in Cycle 0 typically took several months to complete. In full science, we fully expect that the majority of projects will be completed and delivered on an acceptably short timescale, and that relatively few users will feel the necessity to access the raw data, providing very little advantage to “expert” users over less expert users. However, there will always be a small percentage of projects that will be started early but not finished until much later in the season (if at all). If the Early Science policy remains in force, data from such projects would sit in the archive for many months before they are processed and delivered.

Other points in favor of immediate access are: (1) it allows community experts in mm/submm interferometry to identify subtle and complicated problems with that data much earlier; it is always better to get as many expert eyes on the data as early as possible; (2) it provides an incentive for young investigators to learn ALMA one level deeper, helping grow the next generation of mm/submm experts; (3) many projects may have no deliveries before the proposal deadline of the next cycle; raw access would allow better motivated proposals based on early results from the current cycle.

NA therefore advocates that, for full science, PIs be granted access to the raw data as soon as it is archived, without any penalty in the proprietary period.