



Atacama
Large
Millimeter /
submillimeter
Array



Proposal for a Pilot Program for Early Raw Data Delivery

Integrated Science Operations Team

Internal Memo Number: ALMA-SSG-ISOPT-002.A

Status: FINAL

Prepared by:	Organization:	Date:
Anthony Remijan (NA ARC Manager), Liz Humphreys (JAO), Hiroshi Nagai (NAOJ), Martin Zwaan (ESO)	NRAO, NAOJ, ESO, JAO	March 6, 2018
Version	Updates:	Date:
0.1	First Draft	March 6, 2018
0.2	Second Draft – edits to the fraction of programs at each executive, proprietary time and support at each executive	March 8, 2018
1.0	Final draft prepared for the Director’s council. Further edits to the proprietary period and grammatical edits.	March 10, 2018

MEMORANDUM

To: ALMA Director's Council

From: Integrated Science Operations Team

Date: March 6, 2018

Subject: Proposal for a JAO Pilot Program for Raw Data Delivery

As a result of the November ALMA Board Meeting, the JAO was charged to develop a raw data release pilot program in order to evaluate the effectiveness and feasibility of delivering raw data to the Principal Investigators (PIs) and specifically to evaluate the possible impact to the ARC and ARC Nodes. It has been suggested that the delivery of raw data to novice users of ALMA will put undue stress on the work at the ARCs and ARC Nodes for data processing support. As such, the action to develop the pilot program was given to the Integrated Science Operations Team (iSOPT) to evaluate the overall feasibility and to mimic a full raw data release policy as much as possible.

At the iSOPT f2f meeting held on 28 January, 2018, iSOPT laid the initial groundwork for the pilot program and agreed upon the following conditions to run the initial study. The following conditions of the pilot program are given in the list below.

1. The duration of the pilot study will be 12 months starting with Cycle 6. The program will begin after iSOPT is given approval and all the technical and programmatic issues are addressed at each Executive.
2. The pilot program can be terminated at any time upon consensus agreement between iSOPT leadership if due cause is identified. An example of such a cause is given in point 6 below.
3. The program must include proposals/programs from all Executives. Each ARC must participate in the program to test the impact to their communities. Each ARC agrees to the set of observing programs randomly selected by iSOPT to conduct the study (in total about 100 MOUSs). The number of MOUSs for each Executive will be decided by Executive fraction (i.e. ~25 MOUSs for EA, ~37-40 MOUSs for EU and NA, respectively as Chile will not be participating in this program.) The MOUs will be selected from all A- and B-rated projects that are active at the start of the pilot program.
4. The types of observing programs used for the pilot will/should include a random list of PIs. To begin, we will focus on those programs that have a limited set of MOUSs and Executions (~2-3) but cover a wide range of observing space. Table I lists the type of

observing parameters that will be delivered as raw according to the above criteria (if available, e.g. TP MOUSs rarely have just 1 or 2 executions) in order to assess if/when the community needs assistance from the ARCs. For example, users may be able to work easily with single pointing, Band 3 TDM observations but may need more assistance with Band 7 full polarization or Band 8, FDM mosaics. The programs will be randomly selected to also ensure a wide range of PI expertise.

5. No new requirements will be made to the existing software for testing the pilot program. All correspondence about the opportunity for raw data delivery to the PIs will take place through the ALMA Helpdesk. PIs (or their designee in each case) will be contacted by the ARC manager at each Executive informing the PI that their program was selected for the raw data delivery pilot program.
 - a. The ARC Manager at each Executive will contact the PI after the number of QA0 passed executions reaches the expected execution count informing the PI about the details of data downloading.
 - b. The PI will be asked whether or not they request raw data delivery by responding to the Helpdesk Ticket. If the PI responds in the positive, the data will be staged on local ftp (or similar) servers in coordination with their local IT support. This mechanism is already in place at most executives as part of the remote data processing workflow. If the PI responds in the negative, the data will not be staged for the PI. This will give a good metric on how many PIs take advantage of this opportunity.
 - c. Tracking the raw data downloads will be done in coordination with local IT.
 - d. For the pilot study, as in the full program, the proprietary period for that MOUS will be impacted for a PI who decides to download the raw data for their program. The proprietary period for that MOUS will start with the download of the all Execution Blocks (EBs) in the MOUS by the PI. This will be made clear to the PI by the ALMA Helpdesk ticket sent by the ARC manager and will be tracked manually by each Executive. The proprietary period start date will need to be adjusted in the Archive.
6. Support for the raw data delivery, both in the mechanism of downloading the data and help with the data processing, will be done through the ALMA Helpdesk. However, tickets requesting assistance with raw data delivery for this pilot program will be given a lower priority than other submitted tickets although will adhere to the standard operating procedures and service level agreements of the ALMA Helpdesk (i.e. PIs will still get a reply within 48 hours but may not have the issue Resolved in that time.). This is necessary to assess the impact of raw data delivery on available ARC support. Tickets or requests for f2f visits must be given lower weight, as specified in the initial raw data delivery proposal, so that an accurate measure of the operational impact of the program

for each ARC, and for the PI, can be done. During the pilot study, the ARC managers will assess the impact monthly and determine if mitigating action needs to take place if in fact the workload becomes unmanageable. In the extreme case, the pilot program can be cancelled at any time if the workload becomes too excessive.

7. The Integrated Science Team (IST) will assess the overall scientific impact of the program. iSOPT will assess the operations impact of the program. Operations impact will include, but is not limited to, the additional load of the # of HD tickets submitted and will track f2f support for raw data, # hours used for staging the data and the fraction of PIs that actually took advantage of the opportunity. Success will be determined by a measure of the operational load versus the overall scientific impact and overall PI satisfaction through a user survey. Determining the measure of scientific impact and the contents of a user survey is still under discussion.

Any comments or concerns about the program from any Executive, the JAO, Integrated Team or Subsystem can be added as part of this document as an appendix for consideration by ALMA Management and the ALMA Director's Council (See below).

Table 1: Proposed ALMA Datasets to be considered as Part of the Pilot Program

Dataset Description	Band	Correlator Configuration
Single pointing, point-like source, dual polarization	Bands 3 - 7	TDM, FDM
Single pointing, extended source, multiple spatial scales requiring 7-m and TP arrays, dual polarization	Bands 3 – 7	TDM, FDM
Small Mosaics (<50 pointings), dual polarization, 12-m array	Bands 3 – 7	TDM, FDM
Large Mosaics (>50 pointings), dual polarization, 12-m array	Bands 3 – 7	TDM, FDM
Single pointing, point-like source, dual polarization, high frequency	Bands 8 – 10	TDM, FDM
Single pointing, extended source, multiple spatial scales requiring 7-m and TP arrays, high frequency, dual polarization	Bands 8 – 10	TDM, FDM
Single pointing, point-like source, dual polarization, long baselines	Bands 3 – 7	TDM, FDM

Single pointing, extended source, multiple spatial scales requiring 7-m and TP arrays, long baselines, dual polarization	Bands 3 – 7	TDM, FDM
Single pointing, point-like source, full polarization	Bands 3 - 7	TDM, FDM
Single pointing, point-like source, ephemeris object, dual polarization	Bands 3 - 7	TDM, FDM
Single pointing, extended source, multiple spatial scales requiring 7-m and TP arrays, ephemeris object, dual polarization	Bands 3 – 7	TDM, FDM
Single pointing, dual polarization, 7-m array only	Bands 3 – 7	TDM, FDM
Small Mosaics (<50 pointings), dual polarization, 7-m array only	Bands 3 – 7	TDM, FDM
Large Mosaics (>50 pointings), dual polarization, 7-m array only	Bands 3 – 7	TDM, FDM

APPENDIX A: Comments or Concerns from EA for consideration.

APPENDIX B: Comments or Concerns from EU for consideration.

APPENDIX C: Comments or Concerns from NA for consideration.

None

APPENDIX D: Comments or Concerns from the JAO for consideration.