DRAFT: ALMA Archive and User Portal Policy Document

Version: A.1.3
Status: DRAFT

2010-11-29

Prepared By:	Organization	Date
M. Lacy, F. Stoehr, S. Leon Tanne, M. Saito	**ASA working group**	A.1.2: 2010-09-29; A.1.3:2010-11-29
Approved by:	**Organization**	**Date**
Released by IPT Lead(s):	**Organization**	**Date**

1. Introduction

This document presents the policies for handling data and metadata during the proprietary period associated with an ALMA observation, and for user access via the User Portal.

The User Portal is the interface to ALMA for users. It contains tools and capabilities for which an authenticated user identification is required, for example, to track projects, submit helpdesk tickets, or access proprietary data.

The ALMA archive supports several different types of data, grouped into the following categories:

- Metadata: Data entities describing the observations and the bulk data.
- Instrumental data: Visibility data and pipeline images
- Event data: monitor, event and logging data
- Auxiliary data: Other data required to support science operations, including calibrator and spectral line catalogues, documentation and observing proposals.

This document is concerned with the policies for bulk data, metadata and auxiliary data (specifically observing proposals), with an emphasis on the needs and design implications for the ALMA Science Archive (ASA). We have designed these policies to be consistent with AOPvD, the current ALMA Archive Subsystem design document (COMP-70.50.00.00-001-F-DSN) and best practices at other major public observatories such as Gemini, HST and Spitzer.

In three cases we have proposed alternative/additional policies for consideration by reviewers, these are indicted by italics.

2. Policy for metadata

Metadata, in the context of this document, is defined to include everything that describes an observation, with the exception of monitor, logging and event data. This includes all project-related data (with the exception of proposal data, see Section 7) and scheduling blocks. The archive has to persistently store and give access to all metadata. All metadata from all accepted proposals will be made publicly available by default. In cases where metadata are considered sensitive by the PI, the metadata may be nulled or scrambled for the duration of the proprietary period of the program, on a case-by-case basis, with the approval of the Director (or designee). Metadata should be accessible as fast as possible to allow PIs and other external users to monitor the progress of observation programs, and for proposers to check for existing data before proposing new observations, even if the instrumental data itself is still within its proprietary period. If an accepted project is not, for some reason, scheduled on the telescope, its metadata should still be available.

3. Policy for Instrumental Data (visibility data and pipeline images)

The archive must provide means to implement access control to instrumental datasets in order to allow for proprietary periods. The default proprietary period is 12 months, but extensions can be granted upon request to the Director (or designee), and calibration data is publicly available immediately after archiving. Thus the archive must provide support for multiple periods simultaneously, even within a single ASDM file, i.e. there will be some datasets with no proprietary period, many datasets with the 'standard' period and some with an extended period for special projects.

The ObsUnitSet is the minimum set of data on which a pipeline run is triggered. The scope of an ObsUnitSet is defined in Appendix A. The proprietary period is defined as running from 12 months after an ObsUnitSet is completed. This implies that the data has been taken, a pipeline run followed by QA2 has been performed, and the results (image cube and raw data) archived and made available to the PI (Section 4.15). Terminated ObsUnitSets will be treated as completed as of the date of termination.

ALMA is queue scheduled, so a typical ObsUnitSet may take weeks or months to complete, so it must thus be possible to give early data access to PIs before the ObsUnitSet is finished. This will be achieved by allowing PIs access to completed ASDMs as they are archived.

For data which fail QA0 or QA1 due to faults with the observatory hardware or software the observation is reinserted into the queue automatically, The failed data shall be archived, but made available only to ALMA staff. The same policy applies for observations that fail QA2 for reasons associated with telescope hardware or software rather than with the pipeline software.

Duplicate observations should be very rare, as duplication checking is performed at the proposal review stage, both within the set of submitted proposals, and against the database of previously accepted ones. Duplicate proposals may be merged by the review committee, if this is done, PI data access rights extend to the PIs of both the merged proposals. If, despite these safeguards, an observer feels that his/her observation is being duplicated by another approved observation, he/she may request the ALMA Director (or designee) to embargo the dataset in question until the proprietary period on the original dataset has expired.

ALMA staff (including both staff at JAO and the ARCs) shall have access to all data at all times as necessary for technical analysis and tuning of performance. ALMA staff should not use ALMA data for scientific purposes from projects for which they are not PI or CoI until the proprietary period has expired. (based on AOPvD, section 4.15).

Availability of datasets is controlled by access control records, which are part of the metadata information stored in the database. These access control records need to be maintained on a scan-by-scan basis to allow for different proprietary periods for different

sources and extraction of calibration data from ASDMs (see Section 4). Access to datasets in general is only granted to authorized users, but publicly available datasets are available to any authenticated user (including a generic VO user). When metadata is returned in the archive interface, a distinction needs to be made between datasets which are still proprietary and those that are public. This could be achieved by having a "release status" column in the interface that has a release date set far into the future for datasets whose programs are incomplete, and is set to the actual release date once the program is complete and the proprietary period is known.

Alternative: In the interests of ALMA archival research there may be a decision to take more data than requested by a PI. An example would be a project using only one baseband for high-resolution spectroscopy and not using the remaining basebands in continuum mode. If ALMA decides to do e.g. an unbiased observatory continuum survey, the data taken in the additional bands could be made available to the users without a proprietary time. Overlaps with existing proposals and data would need to be checked. If additional data is taken by the Observatory, this will require an extension of the data access and QA records to the baseband level.

4. Calibration data

Calibration data taken in Scheduling Blocks whose sole purpose is calibration (i.e. no science targets) shall be made public immediately. In all cases, fluxes of calibrators recovered by the pipeline shall be entered into the calibrator database for monitoring purposes. In special circumstances, such as a failure of a calibration observation, ALMA staff at the user's ARC will be allowed to extract calibration data from proprietary ASDMs for use by other observers.

5. Virtual Observatory access

The Virtual Observatory (VO) shall be able to access the ASA database using the Table Access Protocol (TAP), and extract observational metadata. In addition, access to non-proprietary image cubes shall be provided following the SIAP version 2 protocol (when that is finalized). Anonymous downloads of data from the ALMA archive must be allowed to enable use by the VO. VO users cannot be required to authenticate through the ALMA user portal to obtain VO accessible data and metadata.

6. Event data

Event data (monitoring and logging) is archived at the OSF and not replicated to SCO and the ARCs. These data shall be accessible to ALMA staff only. If required, the data may be given to observers by ALMA staff in special cases (for example, to investigate a

particular problem with a given dataset), but there will be no general availability of these data to outside users.

7 Policy for proposal auxiliary data

The PI name, proposal code, scientific category, title and abstract of accepted proposals will be made public and accessible through the ASA following approval by the proposal review committee. Ideally, this would include a full text search capability on the abstracts. Science goals, science justifications and technical cases will not be made public, but will be accessible to ALMA staff.
Alternative –I : the proposal data may not be made public until the program is completed (or terminated) with data actually taken. This would avoid the situation where proposal information is made available for low-ranked proposals that were accepted, but for which no observations were actually performed, as the observations were of too low priority in the queue.
Alternative-II: we follow current ESO policy and make only the titles and PIs public immediately.

7. Other (non-proposal) auxiliary data

All non-proposal auxiliary data (calibration database and spectral line databases and observatory documentation) is public without exception.

8. User Portal Policy

Users are required to self-register at the ALMA User Portal. At registration, users need to identify their home ALMA Executive to charge time to (North America (NA), ESO (EU), East Asia (EA),Chile, Other), and their preferred ARC for support (NA,EU,EA). Drop-down menus will allow the following choices for Exec: NA, EA, EU, Taiwan (50%EA, 50%NA), Joint NA/EU, Joint EU/EA, Joint EA/NA.
The policies for user self-identification are as follows:

- Users at institutions in Canada and the US should identify with the NA Exec and request support from the NA ARC.
- Users at institutions in ESO member states (except Chile) should identify with the EU Exec and request support from the EU ARC.
- Users at institutions in Japan and South Korea should identify with the EA Exec and request support from the EA ARC.
- Users at institutions in Taiwan should select the “Taiwan” Exec choice, that charges 50% to EA and 50% to NA.
- Users at institutions in countries not involved in ALMA construction (non ALMA-member states) should select “Other” as their Exec and have a free choice of ARC for support.

- Users at Chilean institutions should select "Chile" as their Exec and have a free choice of ARCS for support.
- JAO staff users, and staff of other non-Chilean telescopes based in Chile (ESO, Gemini etc) should select the Exec corresponding to the organization that pays them (NRAO/NOAO select NA, ESO select EU, NAOJ select EA). They have a free choice of ARC for support.
- Users with joint affiliations split between ARCs should select "Joint XX XX" as appropriate for their Execs, the time will be charged 50% to each Exec. They may choose either ARC for support.
- Users with joint affiliations split between an ALMA member state and a non-ALMA member state should select the Exec and support ARC corresponding to that member state.

User information will be checked for consistency with the above policies after registration (but registration will not be held up for these checks).

Moving institutions

If an observer moves institution they should change their ALMA profile to reflect their new affiliation. Proposals submitted prior to the move will not be affected by the move (for example, if an observer moves from an EA country to an NA country the time for the proposal will still be charged to EA, and their support will be carried out through the EA ARC for that proposal to avoid an ARC having to take on a project for which another ARC was initially responsible).

Password Reset

A mechanism will be provided whereby users who forget their portal password will be automatically be sent a password reset upon request.

To be added/discussed: discussion of Duplicate database entries (unique user identifier=published papers ?), Policy for ARC-dependent branding of UP, Merge with Felix's document.

A. Appendix : Scope of the ObsUnitSet

The ObsUnitSet is the data unit upon completion of which a pipeline run is triggered. The ObsUnitSetStatus is a unique ID for the ObsUnitSet. In practice, the contents of an ObsUnitSet need to be defined, and our suggestion for this is given in this Appendix. In the ALMA data hierarchy, the ObsUnitSet lies above the Scheduling Block (SB), but below the Project, i.e. an ObsUnitSet can contain one or more SBs, and a project can contain one or more ObsUnitSets.

We propose that the contents of an ObsUnitSet be restricted to a set of observations having a common set of calibrators, with a single correlator setup. The simplest case is of a single science target observed in one or more scheduling blocks. More complicated cases might be a mosaic with several pointings in a single field, calibrated off a common phase calibrator, or a single dish raster map. A spectral scan is treated as a single correlator setup for these purposes. Observations of several targets with the same correlator setup would typically be multiple ObsUnitSets unless the objects are close enough in the sky to share a common phase calibrator.

Table 1: Summary

Type of Data	Availability within proprietary period	Availability after expiration of proprietary period

Metadata	All, except by special request.	All
Instrument data passing QA0-2	Only to PI (or co-PI in the case of proposals merged by the proposal review committee)	All
Instrument data failing QA0-1, QA2 for telescope reasons	Only to ALMA staff	Only to ALMA staff
Calibration data	All passing QA0-1	All passing QA0-1
Event data	Only to ALMA staff (may be passed on to observers on case-by-case basis)	Only to ALMA staff (may be passed on to observers on case-by-case basis)
Proposal Title, Abstract, PI name and code	After acceptance of proposal *(alternative: after program completion/termination)*	All
All other proposal data	Only to ALMA staff	Only to ALMA staff
Other (non-proposal) auxiliary data	All	All