

DRAFT: ALMA Archive and Science Portal Policies and Procedures

Version: A.1.17

Status: DRAFT

2012-06-20

Prepared By:	Organization	Date
M. Lacy, F. Stoehr, S. Leon Tanne, M. Saito, T. Wiklind, L-A Nyman	ASA working group	2011-8-25
Approved by:	Organization	Date
Released by IPT Lead(s):	Organization	Date



Table 1: Version history

Version	Date	Affected Sections	Change request numbers added	Comments
A1.5	2010-12-17	All	SOCR-020,021	Revised from original drafts based on comments by JH, FS., version history added
A1.6	2011-01-10	All		Further refinement ; SSO Policy, registration review and duplicate user policies added.
A1.7	2011-01-14	All		Added Science Portal policies (from T. Wiklund), further small changes.
A1.8	2011-01-17	All		Comments from Stephane Menateau (ESO User support lead) and Fernando Comeron (ESO Data Managment lead) added ; subsections relabeled.
A1.9	2011-02-10	All		Appendix A removed (ObsUnitSet définition is responsibility of OT team). SciOps CRs resolved, version sent to DSO for editing
A1.11 (incl A1.10)	2011-02-20	All		Edits by Antonio Hales and Lars-Ake Nyman. Comments from John
A1.12	2011-03-06	All		LN: removed AP05 and AP06 and renumbered. Inserted comments from Doug (ASAC)
A1.13	2011-05-05	7		Changed wording of AP08
A1.14	2011-08-25	10,12,18		Addressed ASAC comments on sections 10,12 Added section 18 on dataset identifier policy.



A1.15	2011-10-21	10.19		Alternative proposal data policy included in section 10 per NRAO request, Section 19 (ingestion of reprocessed data) added.
A1.16	2012-03-02	6,14,18		Science portal update responsibilities clarified, proposal data policy changed (DC decision 28 th Feb 2012), full acknowledgment text provided
A1.17	2012-06-20	6,18		SP responsibilities revised, board-approved acknowledgment text added



1. Introduction

This document presents the policies for the set-up, distribution, evolution and content of the ALMA Science Portal, the standards for Observatory end user documentation, and for the handling of archival ALMA data and metadata during and after the proprietary period associated with an ALMA observation. The Science Portal is a website designed to be the main entry point for the interaction of users with ALMA. It is intended to provide a one-stop access gateway to all ALMA web resources, documents and tools relevant to users for proposal preparation, project assessment, project tracking, project data access, data discovery and retrieval as well as to the ALMA Helpdesk. The Science Portal has components that do not require authentication to access (for example, documentation and tools), and also an authentication mechanism, allowing registered users access to restricted content or tools, depending on the user's role. The authentication mechanism, and website allowing access to the restricted material, was formerly known as the User Portal, whose requirements are listed in the ALMA User Portal requirements document.

Archival research is becoming increasingly important, even from targeted, PI-driven facilities such as HST, Chandra, Spitzer and Gemini. Archival research currently accounts for half of the HST and Chandra papers published each year, including many of the most highly-cited papers (White et al. 2010). It is therefore important that archival research be facilitated by the policies outlined in this document.

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

Metadata: Data entities describing the observations and the bulk data (excluding proposal 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 best practices at other major public observatories such as ESO, Gemini, HST and Spitzer, whilst bearing in mind the policies outlined in the



ALMA Operations Plan version D (AOPvD) and the current ALMA Archive Subsystem design document (COMP-70.50.00.00-001-F-DSN).

Individual policies are identified by APXX identifiers, listed at the top of each subsection and noted in the text in subsections containing multiple policies. Policies are either identified with a section in the AOPvD, the User Portal Requirements (UPR), the Archive Subsystem Design (ASD; COMP-70.50.00.00-001-F-DSN), the Science Software Requirements Document (SSR; ALMA-70.10.00.00-002-N-SPE) or with a SciOps change request (SOCR).

2. Science Portal Access

AP01 [UPR Sect 4.3]: Policy for access to content and tools:

All public (non proprietary) resources including public tools, public web content, public metadata, helpdesk knowledgebase and public science data shall be accessible anonymously.

Authentication and authorization shall be restricted to services and data access that by construction require it. This includes the submission of tickets to the helpdesk, data access of proprietary data as well as tools that are available only as a function of the role of the user e.g. assessor tools, project tracker and Phase 1 Manager.

3. Science Portal Deployment

AP02 [UPR Sect 4.2]: There shall be three instances of the Science Portal publicly accessible: one at each ARC.

Each Science Portal instance includes all the data, metadata and software needed to function independently, including a replica of the user database, document repository (read only), and software tools. (Some of this capability may not be available for Early Science.) This setup is chosen to provide fast data access to the local user communities, spread of the load of web-, database- and data-access as well as to provide redundancy of the content and functionality for maximal uptime of the system as a whole. Each Science Portal instance is connected to local installations of the archive database, bulk store, request handler and user-authentication server. This assures high availability even if one instance should be temporarily unavailable.



AP03 [AOPvD Sect 4.7]: There shall be a single URL pointer to the ARC Science Portal instances (www.almascience.org). From this URL users are automatically redirected to their geographically closest ARC Science Portal. Web links placed on each instance allow the users to switch to a different ARC Science Portal instance e.g. if some portal functionality is temporarily down at one ARC.

AP04: The Science Portal instances at the ARCs must be reachable directly (and bookmarkable in the web-browser) so that the single URL pointer is not a single point of failure. Links to the portal are allowed directly from the websites of the parent organizations, ARCs and ARC nodes.

4. Science Portal User Experience

AP05 [UPR Sect. 4.2]: The instances of the ALMA Science Portal running at the different ARCs should provide the same user-experience to assure ALMA-wide consistency of the user-interactions with the system. This includes the layout, behaviour, content and functionality – the “look and feel”.

In order to provide a clear identification of the local ALMA Science Portal instance with the parent bodies and/or regional funding agencies and to provide local ARC content, minor customizations are allowed:

Logos of the relevant bodies can be placed onto a specified location on the portal

The page title can reflect the local institution

An entry in the main menu for additional content relevant only to the local ARC is possible.

Translations in other languages (e.g. Japanese) can be provided

The content of the ARC menu entry and sub-menu items are the responsibility of the ARCs. They must follow the ALMA Science Portal look and feel but other than that may follow the web policies of the respective ARC.

5. Portal and Helpdesk URLs

AP06: The Science Portal and helpdesk URLs for each ARC shall follow a common scheme. The science portal URL shall be of the form almascience.XXX where XXX is the institutional URL root:



almascience.eso.org, almascience.nrao.edu, almascience.nao.ac.jp. Similarly, the helpdesk shall be of the form help.almascience.org.

6. Science Portal Evolution

AP07: There shall be a single central development instance (e.g. located at JAO) of the Science Portal available where new science content and features are added and tested. Science portal updates may either be full releases, or incremental changes to content. Updates and new releases shall be distributed from the JAO to the ARCs. News items require a 48hr period for comment from the ARCs between being sent out by the JAO and posted to the Portal. Write access to the development version is granted to the ARCs so they can update and maintain their local content and branding directly on the main development version of the Portal. It is envisaged that new functionality and content, if it is relevant for more than one ARC, will be added to the Science Portal instead of the local ARC content. Overall responsibility for Science Portal content resides with the SciOpsIPT.

7. User Registration

AP08 [UPR Sect 4.4]: Users will be required to self-register at a registration page linked to the master User Database at JAO. At registration, users need to identify their home ALMA Region to charge time to (North America [NA], ESO [EU], East Asia [EA], Chile, Other), and their ARC for support (NA,EU,EA).

The policies for user identification are as follows:

Users at institutions in North America will be identified with the NA Region and receive support from the NA ARC.

Users at institutions in ESO member states (except Chile) will be identified with the EU Region and receive support from the EU ARC.

Users at institutions in Japan will be identified with the EA Region and receive support from the EA ARC.

Users who have access to ALMA through more than one region (e.g. due to joint appointments or through official agreements with the Executives) may elect to have their time charged to either region or to have it split between the regions in proportion to their overall share of time. Such users may select either of the relevant ARCs for support.



Users at institutions in countries not involved in the ALMA project (non-ALMA member regions) will have “Other” as their Region and a free choice of ARC for support.

Users at Chilean institutions will have “Chile” as their Region (ARC for Chilean support TBD).

JAO staff users, and staff of other non-Chilean telescopes based in Chile (ESO, Gemini etc) should select the Region corresponding to the organization that pays them (AUI [NRAO], and AURA [LSST/NOAO/Gemini] select NA, ESO select EU, NAOJ select EA).

Users with joint affiliations split between an ALMA member state and a non-ALMA member state will be assigned to 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). Users will be able to change their profiles at any time. The profile to which time will be charged against is the one at the time of proposal submission.

Moving institutions

If an observer moves institution they should change their ALMA profile to reflect their new affiliation. The Exec to which a proposal is charged will not change for proposals submitted prior to the move, but their ARC for support will (for example, if an observer moves from an EA country to an NA country immediately after an ALMA proposal deadline, the time for the proposal will still be charged to EA, but their support will be carried out through the NA ARC).

Password Reset

A mechanism will be provided whereby users who forget their portal password will be automatically be sent a password reset upon request. There will also be a mechanism to recover forgotten userids based on the user’s name or other uniquely identifying information (PI project code etc).

Duplicate entries in the user database

Duplicate users database entries are likely to occur from time to time, for example if a user moves institute and re-registers rather than changing their profile, or if a user forgets their user identifier, or that they are already registered. (A check may be made for duplicated userids upon registration, but this may not always be effective if a user picks a different userid.) Currently, there is no recognized international



unique and persistent identifier for astronomers, though it may be possible in future to base one on, for example, a publication record in the major journals. Until that time, duplicate entries will need to be weeded out by hand. Once a duplicate entry is detected, the user will be sent an email (to both registered email addresses if appropriate) and asked which userid should be retained. The others will be merged.

Registration reviews

Registration reviews will be carried out periodically, jointly by JAO and ARC staff to ensure compliance with the registration policies set out in this document. If a users profile is found to be at variance with these policies, an email will be sent to the user requesting that they update and correct their profile.

8. Single Sign-On [UPR Sect 4.3]

AP09 [UPR Sect 4.3]: The ARCs are encouraged to provide single-sign-on systems so that registered and authenticated users of the local ARC portals are automatically registered and authenticated at the ARC instances of the ALMA Science Portal. The ARCs are also encouraged to automatically register their users with ALMA. (These latter two capabilities require additional development work before they are implemented.)

9. End User Documentation standards

AP10: Document Template:

The document template for ALMA end-user documents shall conform to the standard ALMA end-user format. This format is defined through the ALMA end-user document template, maintained by the JAO.

The ALMA end-user document template consists of three pages:

Page 1 (front)

Document number and version number

Title

Author (in special cases)



ALMA logo

ARC logos

Page 1 (back)

Direction for User Support in the form of URL(s).

Comments regarding the document

Revision history with dates and authors/editors

Space for acknowledgment of external contributors

Space for logos of external contributors

Format for referencing the document

Last page

ALMA logo

Acknowledgment of Executives and funding agencies

Executive logos

In addition a standard format will be defined for chapter headings, footnotes, headers and figure captions. The ALMA end-user document template is in US Letter format.

AP11: Document Revision Control System:

The versioning of the ALMA end-user documents shall be done through a revision control system (RCS) (implementation most likely through Plone). Documents shall be stored in a repository containing all versions of a given document, physically located at the JAO office in Santiago. Authors and editors can download a document from the repository for local editing. This check-out process will lock the document and no other access is possible until the author/editor check-in the document. Once the document is checked-in it will automatically be given a new version number. A Content Manager at the JAO DSO will oversee the process and keep track of which documents that are checked out at any given time. An author/editor shall notify the Content Manager whenever a document is checked out and in via email.

In order to be eligible to check documents in and out of the repository the author/editor needs to be registered within the Revision Control System. The registration is done via the DSO Content Manager.



The registered author/editor shall be given a user role (Editor) and a password to enable access to the document repository.

AP12: End-user Document Verification and Acceptance:

The new or revised document shall be written in Word, LaTeX or Pages format but all documents published on the web pages will be in pdf format. Earlier versions of a document are stored in the original Word, LaTeX or Pages format. Hence, only publishable documents are retained in pdf format.

Before an end-user document or a revised version of it gets published on the ALMA Science Web pages it needs to go through a verification and acceptance process.

A Content Manager at the JAO DSO is responsible for ensuring that new or revised documents are transformed into pdf format and published on the ALMA Science Web pages. This procedure involves three steps: (1) verification of the document with regards to content and format, (2) acceptance of the content and format, and (3) transforming the document into pdf format and publishing on the web pages.

When a new document or a revised version of a document is uploaded to the document repository, the author/editor shall notify the DSO Content Manager via email. The Content Manager is then responsible for ensuring that the document is verified, accepted and published on the ALMA Science Web pages.

The goal of the verification process is to ensure that the document is consistent with existing ALMA capabilities, policies and procedures. The Content Manager can involve other DSO and/or ARC staff in this process. If any inconsistencies are found the Content Manager will contact the author/editor of the document in question to resolve the inconsistency. The Content Manager shall also ensure that the format of the document is consistent with the ALMA end-user document template and format.

The acceptance of the verified document is the responsibility of the Head of DSO. The Content Manager will present the new or revised document to the Head of DSO with a summary of the pertinent issues and/or revisions. In most cases this will be sufficient for approval but cases involving statements and/or changes involving ALMA policies the Head of DSO may deem it necessary to get approval from the JAO Director, Directors Council or the ALMA Board. The decision to request approval for a given end-user document is the responsibility of the Head of DSO.



Once the Head of DSO has accepted a new or revised document the Content Manager shall ensure that it gets transformed into pdf format and published on the ALMA Science Web pages. If the document replaces an earlier version the old pdf file shall be removed from the repository.

10. Archive policy for metadata

AP13: Policy for metadata: Metadata from all observations will be made available without restriction upon archiving of the data.

Metadata, in the context of this document, is defined to include everything that describes an observation, with the exception of proposal, monitor, logging and event data. This includes all project-related data and scheduling blocks. The archive has to persistently store and give access to all metadata [ASD 4.4-R2].

Metadata from all observations will be made publicly available upon archiving of the data. 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, executed on the telescope, no metadata or proposal auxiliary data shall be made available (see also AP18).

11. Archive Policies for Instrumental Data (visibility data and pipeline images)

AP14: [AOPvD sect 4.15]. 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 variations on this can be granted upon request to the Director (or designee) Thus the archive must provide support for multiple periods, i.e. there will be some datasets with no proprietary period, many datasets with the 'standard' period and some with a shorter or extended period for special projects.

AP15: [SOCR-021]: The proprietary period is defined as running from 12 months after an ObsUnitSet is completed and made available to the PI (see detailed discussion in AP29). The ObsUnitSet is the minimum set of data on which a pipeline run is triggered, and lies between the Project and individual Scheduling Blocks in the ALMA data hierarchy. In practice, the scope of the ObsUnitSet



will be implemented in the OT. Terminated ObsUnitSets will be treated as completed as of the date of termination.

AP16: [SSR 7.1-R1.1]: All ALMA data shall be archived by default, regardless of its quality. For data that 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 made available only to ALMA staff.

AP17: The same applies [AP18] for observations that fail QA2 for reasons associated with telescope hardware or software rather than with the pipeline software.

AP18: duplicate observations: 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. 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. No duplication checks will be made against Cycle 0 or Cycle 1 observations due to the expectation that developing ALMA capabilities will rapidly outstrip those available at the start of Early Science.

AP19: [AOPvD, section 4.15]: 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 shall not use any ALMA data for scientific purposes from projects for which they are not PI or CoI until the proprietary period has expired.

AP20: [SOCR-020, approved (also agreed at ALMA management meeting in March 2008 per B. Glendenning)]: Publicly released data shall have anonymous access: Availability of datasets is controlled by access control records, which are part of the metadata information stored in the database. These will be set at the ObsUnitSet level. Access to datasets in general is only granted to authorized users, but publicly available datasets are available to any user anonymously (variance from ASD sections 9.1.7 and 9.3, but now approved). When metadata is returned in the archive interface, a distinction shall be made between datasets which are still proprietary and those that are not. *{Implementation detail: this will be achieved by having a “release status” column in the interface that has a release date set far into the*



future in the time between when the data are placed in the archive until the ARC staff confirm that the data is available to the PI and set the actual release date, AP27}

AP21: Additional data: In the future, 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 spectral window for high-resolution spectroscopy and not using the remaining spectral windows in continuum mode. If ALMA decides to do e.g. an unbiased observatory continuum survey, the data taken in the additional spectral windows will become available only after the proprietary period on the rest of the data has expired. Overlaps with existing proposals and data will be checked. (Note: in principle, these data could be made non-proprietary immediately, but this would require extending the access control records to the spectral window level, and ASDMs would need to be split to separate the proprietary and non-proprietary data, requiring significant implementation effort. We believe such effort would be unjustified in the short term, although this could be revisited at a future date.)

12. Archive Policy for Calibration data

AP22: [SOCR023 (approved)]: Calibration data taken in ObsUnitSets whose sole purpose is calibration (i.e. no PI-specified science targets) shall have no proprietary period, and be available immediately upon archiving, without restriction, following successful QA2. Results from non-science target observations included within a PI science project may be extracted by the JAO and incorporated into observatory-maintained tools such as calibrator databases. These metadata will have no proprietary period. In special circumstances, such as a failure of a calibration observation, ALMA staff at the user's ARC shall be allowed to extract non-science target calibration data from proprietary ASDMs for use by other observers. Such activity would be generated by a helpdesk ticket, and carried out at the discretion of the ARCs. Note: this represents a variance from AOPvD Sect 4.15.

13. Archive Policy for Event Data

AP23: Event data shall be accessible to ALMA staff only: (system 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.

14. Archive Policy for Proposal Auxiliary Data

AP24: For proposals that are ranked so highly that they are expected to be observed during a given ALMA observing cycle (grades A+B), the following information will be made public soon after PIs are informed of the outcome of the assessment process:

Project number and Title

Name of PI, region of PI

Names of CoIs

For filler projects (C), the corresponding information will be made public as soon as the first ObsUnitSet/dataset is archived.

Proposal Abstracts are released only when data proprietary period ends

15. Archive Policy for Other (non-proposal) Auxiliary Data

AP25: All non-proposal auxiliary data (calibration database and spectral line databases and observatory documentation) is public without exception.

16. Archive Policy for Virtual Observatory Access

AP26: [SOCR-020]: VO users shall access the ALMA Archive anonymously: 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 [SOCR-020].

17. Data delivery



AP27: [AP20], [SOCR-024 (approved)]: Data delivery: the proprietary time will begin when the PI is notified by email: As soon as the pipeline-processed, QA2 validated ObsUnitSet is received by the home ARC, and checked for corruption (via a checksum or similar) the PI will be notified by email, and the proprietary period will begin. These actions will be triggered by a process run at the home ARC, which will enter a state into the master ASA database at the SCO, which will then set the release date in the ASA database.

PIs will be notified via email by the ARCs following the successful transfer of an ObsUnitSet to the relevant ARC Mirror archive. From this point on, the processed datacubes (either processed by JAO staff during Early Science, or by the pipeline later on) will be available to PIs via the internet. Raw data from completed ObsUnitSets will be made available to the PIs upon request. In general, downloads over the internet will be preferred. Media shipment of data to registered users will be possible, following the policy of the relevant ARC [TBD whether we can agree on a common policy]. After the proprietary period, anonymous data transfer over the internet will be allowed (SOCR-020).

Note variance with AOPvD Section 4.13, which specifies that users will automatically be sent a data package upon the completion of their program.

18.Dataset identifiers

AP28: Dataset identifiers: ALMA datasets will be identified in papers through the project ID, in the acknowledgements.

The following statement should be included in the acknowledgment of papers that make use of ALMA data:

“This paper makes use of the following ALMA data: ADS/JAO.ALMA#<Project code>. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada) and NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.”

19.Ingestion of reprocessed data into the archive

AP29 Data produced at the ARCs as part of a project-sanctioned reprocessing shall be returned to the JAO, where QA2 will be performed and the data ingested into the archive.



[wording still TBD] It is possible that at some point an improved pipeline will be produced that will result in better data products. In the event that the cluster at SCO is too oversubscribed with processing new data to be able to undertake reprocessing of archival data, that reprocessing may be performed at the ARCs upon the request of the JAO data management team. When such products are produced, they will be transferred to the JAO from the responsible ARC, and QA2 will be performed. Assuming this QA2 is successful, the products will be ingested into the ASA and replicated to all three ARCs.

AP30 Data reprocessed without a request from the JAO will not be placed into the ASA.

[Exact wording TBD] Some users may wish to return their fully reprocessed ALMA data to their ARCs for archiving as part of their data management strategy. Such data will not be ingested into the ASA, but may be ingested into institutional archives (NRAO, ESO or NAOJ) and made available to ASA users through Virtual Observatory protocols.

References:

White, R.L. et al. 2010, Astro2010: The Astronomy and Astrophysics Decadal Survey, Position Papers, no. 64

Table 2: Summary

Type of Data	Availability within proprietary period	Availability after expiration of proprietary period
Metadata	All (for executed observations)	All (for executed observations)
Instrument data passing QA0-2	Only to PI or co-PI	All
Instrument data failing QA0-1, QA2 for telescope reasons	Only to ALMA staff	Only to ALMA staff
Calibration data	All ObsUnitSet that consist of calibration data only passing QA2; calibrator fluxes (in calibration database), other data	All passing QA2



	on case-by-case basis.	
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, PI name and code	All A&B graded proposal, grade C on archiving of first ObsUnitSet	All
Proposal abstract	Only to ALMA staff	All
All other proposal data	Only to ALMA staff	Only to ALMA staff
Other (non-proposal) auxiliary data	All	All