Atacama Large Millimeter Array	ALMA-SW-666 Revision:3.0 2001-07-24 *Requirements* Steven T. Myers

ALMA Offline Data Processing Requirements

Requirements

ALMA Pipeline and Offline Software Group:
B. Clark, F. Gueth, M. Momose, S. Myers,
P. Schilke, K. Tatematsu
ALMA Software Science Requirements Committee:
H. Ezawa, T. Handa, R. Lucas, J. Mangum, T. Ohnishi,
J. Richer, S. Scott, F. Viallefond, M. Wright
ALMA Software Use Case Group:
G. Harris, D. Muders, J. Schwarz, R. Warmels

Keywords: Requirements, Offline, Science, Calibration

Author Signature: Steven T. Myers	Date: 2001-07-24
Approved by: B. Glendenning, G. Raffi Institute: NRAO, ESO	Signature: Date:
Released by: Institute:	Signature: Date:

Change Record

Revision	Date	Author	Section/ Page affected	Remarks
1	2001-03-06	S.T. Myers	All	Initial Version
1.2	2001-05-01	S.T. Myers	All	Version for SSR Telecon
2	2001-06-21	S.T. Myers	All	Version to SSR
2.2	2001-07-12	S.T. Myers	All	Comments from SSR
3	2001-07-24	S.T. Myers	All	After Berkeley, Remove Pipeline

`$Id: myers-report-2.tex,v3 2001/07/24 myers Exp myers $`

Contents

1 Introduction **5**

1.1 Nomenclature . . . 5

1.2 General Considerations . . . 6

2 Offline Data Processing Requirements **7**

2.1 General Requirements and Interaction with Other ALMA Elements . . . 7

2.1.1 Goals of the Offline Package and Relation to the Pipeline . . . 7

2.1.2 Operational Issues . . . 7

2.2 Interface . . . 8

2.2.1 General User Interface Requirements . . . 8

2.2.2 Graphical User Interface (GUI) . . . 8

2.2.3 Command Line Interface (CLI) . . . 9

2.2.4 Interface Programming, Parameter Passing and Feedback . . . 10

2.2.5 Documentation and Help Facility . . . 10

2.3 Data Handling . . . 11

2.3.1 General Data Requirements . . . 11

2.3.2 Data Import and Export . . . 13

2.3.3 I/O Speed and Efficiency . . . 13

2.3.4 ALMA Interferometer Data . . . 13

2.3.5 ALMA Single Dish Data . . . 14

2.3.6 Images and Other Data Products . . . 14

2.3.7 Foreign Data . . . 14

2.3.8 Interaction with the Archive . . . 14

2.4 Calibration and Editing . . . 15

2.4.1 General Calibration and Editing Requirements . . . 15

2.4.2 Interferometer Data . . . 16

2.4.3 Single Dish Data . . . 17

2.4.4 Mosaicing Considerations . . . 17

2.4.5 Ancillary and Diagnostic Data . . . 18

2.5 Imaging . . . 18

2.5.1 General Imaging Requirements . . . 18

2.5.2 Interferometer Imaging . . . 18

2.5.3 Single Dish Imaging . . . 19

2.5.4 Mosaicing Considerations . . . 19

2.5.5 ACA Imaging Considerations . . . 19
2.6 Analysis . . . 19
2.6.1 General Analysis Requirements . . . 19
2.6.2 Visibility Data Analysis . . . 19
2.6.3 Single Dish Data Analysis . . . 19
2.6.4 Mosaicing and Combined Array Analysis . . . 19
2.6.5 Image Analysis and Manipulation . . . 19
2.7 Visualization . . . 20
2.7.1 General Visualization and Plotting Requirements . . . 20
2.7.2 Display Appearance and Interactivity . . . 20
2.7.3 Image-cube Manipulation . . . 21
2.8 Special Features . . . 21
2.8.1 Simulation . . . 21
2.8.2 VLBI . . . 21
2.8.3 Solar Observing . . . 21
2.8.4 Pulsar Observing . . . 21

1 Introduction

The history of this document and correspondence can be found at:

`http://www.aoc.nrao.edu/ smyers/alma/offline-req/`

This document is `myers-report-3.tex`.

This document describes the requirements for offline data processing software packages in order to be able to handle the ALMA data output. It is assumed that there will be (at least) one software package available for users to reduce their own data and/or data from the archive offline at their home institution, at an ALMA regional center, or remotely using ALMA center computing. This suite of tools must fulfill the Offline Data Processing Requirements (Section 2). We refer to the ensemble that fulfills these as the "Package". Again this might be an assortment of different programs from different software packages (e.g. AIPS, GILDAS, MIRIAD, aips++) but in this case it is highly desirable that there be at least one single package that fulfills the Offline requirements or that there be an installation that integrated the necessary parts. Note, however, that is is highly unlikely that disparate applications from different packages will fulfill the requirements on similar "look and feel" and inter-connectivity, and thus it is likely as well as desirable that this will be a single homogeneous suite.

This document assumes requirements already delineated in ALMA-SW MEMO 11 "ALMA Software Science Requirements and Use Cases" at:

`http://www.alma.nrao.edu/development/computing/docs/joint/0011/ssranduc.pdf`

Much of the content of this document is based on the AIPS++ User specifications Memo 115 found at:

`http://aips2.nrao.edu/stable/docs/specs/specs.html`

Although it was intended to be the basis for AIPS++, these are an excellent starting point for our requirements to build upon. We do not think this makes this too specific to AIPS++, as our document is intended to be package-independent.

1.1 Nomenclature

The subject of Section 2 on Offline Data Processing Requirements is referred to as the *Package* or *Offline Package*. This is intended as a set of tools or programs, believed adequate for ALMA reductions, and used by observers for science and by ALMA staff for reductions upon which the behavior of the system will be judged. It may consist of packages provided by different groups, with transitions provided to integrate them into a single suite. It may be that more than one Package fulfills the requirements of Section 2 and thus can be considered as suitable for ALMA processing. The requirements will state that the Package will be available for installation on the observer's own computer systems as well as present at ALMA centers.

There is another set of requirements, given in ALMA-SW MEMO 11, for the software that forms the ALMA Online Data Processing Pipeline, or simply the *Pipeline*. Note that the Pipeline may or may not be based on the offline Package(s), depending on implementation. However, the functionality of the Pipeline must be available in the offline Package for users.

The user of the package may be referred to as *user* or *observer*, and may be the actual proposer or a staff member. In automated modes, the user may actually be another tool or program.

The *Archive* refers to the totality of the ALMA data storage and consists of possible different physical archives.

The definitions of various terms, such as *scans*, are given in ALMA-SW MEMO 11.

1.2 General Considerations

A. Two fundamentally new aspects of ALMA are the integrated Archive and the Pipeline, therefore the impact of requirements on these two areas should be considered. We have included topics "Relation to the Pipeline" (§2.1.1) and "Interaction with Archive" (§2.3.8).

B. There is a fundamental difference between running the ALMA array in interferometric and single-dish modes. This difference may not be so fundamental for many types of data processing, however. For example, if several single dishes do single-dish together, much of the calibration will be done interferometrically (pointing, focus, beam shape); for interferometric observations, temperature scale is derived from single-dish measurements. Therefore where appropriate we split these two paths, with inclusion of integration of the two where appropriate.

C. There is no fundamental difference between spectral-line and continuum observations, merely number of channels and bandwidths coming out of the correlator. Due to the nature of the ALMA correlator, most of the special calibration needs of traditional spectral line observations (e.g. bandpass calibration) are also applicable to continuum observations (continuum is built up through summation of spectral channels taken at low resolution). We will assume that all data will be effectively taken in spectral line mode.

D. There is no fundamental distinction needed for polarization data, merely consideration of the number of polarization products needed for processing. There may be special modes where polarization (e.g. RR or LL from X and Y) is synthesized in hardware, but these can be treated as special cases. The model we consider here is the processing of one or more Stokes parameters and thus polarization is integrated into all the topics.

E. Mosaicing will likely be widespread in ALMA data reduction due to the small fields of view. In this trial outline, we choose to include mosaicing as cases (like spectral line) under the interferometric and single-dish modes when the requirements are straightforward, but also under special headings "Mosaicing considerations" where it seems more appropriate (under the Calibration and Imaging headings).

F. We propose a system of prioritizing with codes:

1 = essential
2 = highly desirable
3 = desirable, but not critical

These codes are enclosed in brackets [] at the end of the items they qualify.

It is intended that Priority 1 items must be present in the Package and work with high efficiency. Priority 2 items should be in the package, though there may have to be sacrifices in performance or availability may be delayed. Fulfilling Priority 2 items well should give a boost to a given Package's rating. We expect that the Package will fulfill all Priority 1 and 90% or more of Priority 2 requirements. Priority 3 items should be considered for upgrades or development.

G. Specific requirements can/should be broken into sub-cases or instances (e.g. R1 R1.1, R1.2) for clarity. These sub-requirements are either special cases of the main requirement (often with different priority) or itemized instances or examples related to the requirement.

H. There are a number of instances where certain specifications, such as data formats supported, are designated as being in a list maintained by the ALMA project. There are other points in the requirements that refer to "standard modes", which implies a detailed list (tbd) also maintained by the project. These cases are indicated in **boldface** to aid in determining which lists will need to be constructed.

I. At the time this document was written, the ACA has not been officially scoped as part of ALMA. However, some placeholder ACA-related requirements are included in the main requirements section, and are clearly labeled as such.

2 Offline Data Processing Requirements

2.1 General Requirements and Interaction with Other ALMA Elements

2.1.1 Goals of the Offline Package and Relation to the Pipeline

Note: An ALMA Offline Data Processing Package (or "the Package") is primarily intended to enable end-users of ALMA (e.g. observers or Archive users) to produce scientifically viable results that involve ALMA data products. The secondary use is to enable ALMA staff to assess the state of the array and derive calibration parameters for the system, although it is anticipated that the Pipeline will be the primary engine for Online Data Processing and system monitoring.

Also note that not all offline analysis tools will necessarily be in the Pipeline package. For example, one of the important differences between Pipeline and Offline reduction path is that Offline one should have extensive interactive capabilities to merge and compare data with different resolution, coordinate system, data grid, and so on.

2–1.1–R1 There must be an Offline Data Processing Package that fulfills the requirements laid forth in this document.

Priority: 1

2–1.1–R2 All standard observing modes supported by ALMA must be processable by the Package.

Priority: 1

2–1.1–R3 The package should be able to function (be installed) at the users home institution and at ALMA regional centers (both locally and remotely). It should be portable to a reasonable number of supported platforms, including systems without network connections and laptops.

Priority: 1

2–1.1–R4 The performance of the package should be quantifiable and commensurate with the data processing requirements of ALMA output at a given time. This should be benchmarked (e.g. "AIPSmarks") and reproduce accurately results for a fiducial set of reduction tasks on specified test data.

Priority: 1

2–1.1–R5 All standard processing functionality available in the Pipeline must be available in the Package also as an offline analysis option.

Priority: 2

2.1.2 Operational Issues

2–1.2–R1 Installation of package must be flexible, and able to be installed on non-specialized hardware by end user, preferably without root user permission (on Unix).

Priority: 1

2–1.2–R2 User-understandable and non-destructive error handling at all levels is highly desirable. There must be provision for job control such as interrupt and abort.

Priority: 1

2–1.2–R3 The Package providers must provide support for the package, with bug-fixing on timescales approprite to the severity level of the defect (e.g. 1 week or less for catastrophic bugs with no work-around). **These defect levels and timescales should be delineated by the ALMA project**.

Priority: 1

2–1.2–R4 Updates and new versions of the package should be backward compatible where possible, such that user-built and observatory provided scripts and tools should be executable with only minor changes.

Priority: 1

2–1.2–R5 Source code for the astronomical routines in the Package should be available to the user.

Priority: 1

Notes: This does allow for the use of some proprietary data handling, output formatting, and special processing (e.g. Pixon) routines.

2–1.2–R6 User installation of the package should not be restricted by other issues such as expensive or unduly restrictive licenses. The Package license should convey all other necessary licenses (such as GNU) or they should be available to the user for only an nominal fee.

Priority: 2

Comment: This allows for the use of commercial components. Although it may be attractive to build upon a commercial package such as IDL, this is likely to be prohibitive unless a blanket license would be bought by the project and distributable free of charge to users.

2–1.2–R7 There should be the provision for the development and incorporation of user-supplied routines.

Priority: 2

2–1.2–R8 The application of successive stages of calibration, correction, flagging and editing should not be destructive to the data. The package should be able to recover and revert to earlier stages without resorting to saving of the entire dataset at intermediate states.

Priority: 2

2.2 Interface

2.2.1 General User Interface Requirements

2–2.1–R1 User must be able to choose from a variety of interface styles, including:

- **2–2.1–R1.1** A Command Line Interface (CLI) must be provided, with access via both an interactive input and via script.
  Priority: 1
- **2–2.1–R1.2** A Graphical User Interface (GUI) must be provided for interactive processing. Actions taken under the GUI must be loggable, editable, and executable by the CLI.
  Priority: 1

2–2.1–R2 The user should be able to interact with the host operating system with command sequences invoked from the UI.

Priority: 1

2–2.1–R3 Multitasking for all interfaces should be available where appropriate. It must be possible to run one or more long-running calculations in the background. While background tasks are running normal interactive activities must be possible.

Priority: 1

2.2.2 Graphical User Interface (GUI)

Note: The GUI is intended to be the primary interactive mode for users, especially neophyte users, and thus should be tailored for clarity and ease of use. Use of the Package GUIs should be a pleasurable, not a frustrating, experience!

2–2.2–R1 The GUI should provide real-time feedback via standard compact displays:

- **2–2.2–R1.1** Window updating should be fast (¡0.1s on same host).
  Priority: 1
- **2–2.2–R1.2** Windows should not take up excessive screen space, with full GUI controls visible on a window one-third the size of a standard view surface (or approximately 800x800).
  Priority: 1
- **2–2.2–R1.3** Users should have the choice of cascading windows or re-use of single window for new operations.
  Priority: 2
- **2–2.2–R1.4** It is desirable that there be a master control GUI for process control which keeps track of sub-windows and tools.
  Priority: 3

2–2.2–R2 It must be easy to run GUIs remotely from the host machine (e.g. via X displays).

Priority: 1

2–2.2–R3 GUI-based tools should be available for reduction of data taken in all standard ALMA observing modes.

Priority: 2

2–2.2–R4 The default look and feel of the GUI should be uniform and familiar through the entire package.

Priority: 2

2–2.2–R5 The look and feel of the GUI must be customizable to accommodate both the expert specialist and the novice user, with ability to hide complexity when prudent and the ability to access deeper levels when desired. The default look to the novice should not be overly busy, with functionality easily apparent through labeling and built-in help facility.

Priority: 2

2–2.2–R6 It is desirable that it be easy for users to develop and include their own custom GUIs in the Pacakge.

Priority: 3

2.2.3 Command Line Interface (CLI)

Note: the CLI is the primary mode for automatic reduction for ALMA, and it is anticipated that there will be a suite of "standard" scripts developed to help users in data reduction tasks. Thus it is important that the Package support all of its critical modes in the CLI.

2–2.3–R1 The interface must have the facility to read in command files for batch processing of a sequence of CLI commands.

Priority: 1

2–2.3–R2 The CLI should have command-line recall and editing, and name completion where appropriate.

Priority: 1

2–2.3–R3 All functionality of the GUI must also be available in CLI mode (although posssibly with loss of simplicity in instances where the graphical selection is important).

Priority: 1

2–2.3–R4 The CLI should be usable remotely over low-speed modem lines or network connections, with ASCII terminal emulation.

Priority: 2

2–2.3–R5 A CLI mode menu for display and editing of input parameters and execution of tasks is desirable.

Priority: 3

2.2.4 Interface Programming, Parameter Passing and Feedback

2–2.4–R1 The UI must have basic programming facilities such as:

2–2.4–R1.1 variable assignment and evaluation
Priority: 1
2–2.4–R1.2 conditional statements
Priority: 1
2–2.4–R1.3 control loops
Priority: 1
2–2.4–R1.4 string manipulation
Priority: 1
2–2.4–R1.5 user-defined functions and procedures
Priority: 1
2–2.4–R1.6 standard mathematical operations and functions
Priority: 1
2–2.4–R1.7 efficient vector and matrix operations
Priority: 2
2–2.4–R1.8 user-defined data structures
Priority: 2

2–2.4–R2 Commands executed should be logged, with provision to re-execute the session.

Priority: 1

2–2.4–R3 Input parameter and syntax checking should be effected upon function calling or parsing with reporting of incorrect, suspicious or dangerous choices before execution where possible.

Priority: 1

2–2.4–R4 Parameters should be passable between applications in as transparent a manner as possible. However, global variables should not be the default, unless designated specifically by the user-programmer.

Priority: 1

2–2.4–R5 Application variables should be named consistently and as clearly as possible indicating their intended use using astronomical terms where appropriate.

Priority: 2

2.2.5 Documentation and Help Facility

2–2.5–R1 The package creators must provide up-to-date, comprehensive and user comprehensible documentation for all parts of the package.

Priority: 1

2–2.5–R2 There should be a variety of help levels and documentation formats accessible from the UI and over the Internet, applicable to novices, experts, and technical users. These would include:

2–2.5–R2.1 user cookbooks with extensive examples
Priority: 1

2–2.5–R2.2 application descriptions and reference manual (with all inputs to functions and tools)
Priority: 1
2–2.5–R2.3 online help, FAQ, email contacts
Priority: 1
2–2.5–R2.4 release history, bug reports and tracking, patch descriptions
Priority: 1
2–2.5–R2.5 programmer references and guides
Priority: 2
2–2.5–R2.6 data format descriptions
Priority: 2
2–2.5–R2.7 algorithm descriptions
Priority: 2
2–2.5–R2.8 newsletters, email exploders, notes series
Priority: 3
Note: these would be maintained by the package providers, with help from the ALMA project.

2–2.5–R3 Help materials should also be available in printable formats (such as pdf and printer-friendly versions of HTML pages).

Priority: 1

2–2.5–R4 Help should be context-sensitive where relevant. In GUI mode, fly-over banners should indicate use of buttons and fields, and clickable help buttons should be available on all pages.

Priority: 1

2–2.5–R5 In GUI mode, help functions may direct user to Web page. In CLI mode, must support in-line text based help also.

Priority: 1

2–2.5–R6 Full search capability must be built into documentation library.

Priority: 2

2.3 Data Handling

2.3.1 General Data Requirements

2–3.1–R1 The package must support data taken in any of the standard ALMA hardware modes.

Priority: 1

2–3.1–R2 Must be able to handle the integrated data objects corresponding to the observational programs carried out by ALMA. These objects may be implemented in any manner appropriate, though relations between the components of the object must be maintained through some mechanism. These include:

2–3.1–R2.1 Program header information
2–3.1–R2.2 Observation status information (and schedules themselves)
2–3.1–R2.3 Field information
2–3.1–R2.4 Coherence function (visibility) data from interferometer in all available polarization products, spectral channels, frequency bands, IFs, including auto-correlations, in uncorrected and/or corrected form (online WVR correction)
2–3.1–R2.5 Auto-correlations in single-dish total power modes
2–3.1–R2.6 Weights and/or data uncertainties
2–3.1–R2.7 Frequency switching mode
2–3.1–R2.8 Flagging data or masks

2–3.1–R2.9 Diagnostic data and errors
2–3.1–R2.10 A-Priori calibration data (bandpasses, flux densities, polarization leakages, etc.)
2–3.1–R2.11 Derived calibration data (gain tables, flux bootstraps, etc.).
2–3.1–R2.12 Images and/or models produced from data
2–3.1–R2.13 Processing history

Priority: 1

2–3.1–R3 There must be a selection mechanism integrated within tools to choose between the various available data subsets such as:

2–3.1–R3.1 polarization products
2–3.1–R3.2 bands (frequency bands, IFs)
2–3.1–R3.3 spectral channels
2–3.1–R3.4 frequency-switched data
2–3.1–R3.5 interferometer: subarrays
2–3.1–R3.6 interferometer: WVR-corrected or uncorrected baselines

Priority: 1

2–3.1–R4 Multiple pointing centers for mosaics must be supported.

Priority: 1

2–3.1–R5 For polarization products, transformation must be provided to the desired Stokes output parameter(s).

Priority: 1

2–3.1–R6 Averaging of bands and spectral channels (compression) should be possible.

Priority: 1

2–3.1–R7 Data taken in arbitrary (but parameterized) scanning patterns must be dealt with.

Priority: 1

2–3.1–R8 The flagging mask must be maintained and associated with the data it refers to during any subsequent operations (such as splitting of data sets).

Priority: 1

2–3.1–R9 Calibration and ancillary monitoring data must be preserved, if requested.

Priority: 1

2–3.1–R10 Comprehensive and understandable processing history information for the data must be maintained and be exportable (both as tables and as plain text).

Priority: 1

2–3.1–R11 Users should have access (at the manipulation level) to all aspects of the data including the header.

Priority: 1

2–3.1–R12 The Package must support locking data files so that there is no possibility of one process corrupting a file that is also being written to by another process in the Package. The default model should be: "one writer, multiple readers."

Priority: 1

2–3.1–R13 Distinctions between "single-source", "multi-source", single-dish, and interferometer datasets should be avoided with context built into the dataset or header.

Priority: 2

2–3.1–R14 Concatenation of datasets should be straightforward and robust. Extraction and reinsertion of data subsets should be supported.

Priority: 2

2–3.1–R15 When sorting or indexing is desirable for performance enhancement, it is desirable that this be carried out in a manner transparent to the user.

Priority: 3

2.3.2 Data Import and Export

2–3.2–R1 ALMA standard data formats (e.g. ALMA raw data format, FITS) must be supported for both input and output without loss of functionality or information.

Priority: 1

2–3.2–R1.1 Other standard data formats must be supported, possibly with some loss in information content. **The project will maintain a list of formats which the package must support.** *Priority: 1*

2–3.2–R2 Disk and offline data storage (e.g. DAT, DDS, DLT) must be supported. **The project will maintain a list of media which the package must support.**

Priority: 1

2–3.2–R3 The ability to drop flagged data on export should be included.

Priority: 2

2–3.2–R4 The Package internal data format, which may be independent of other supported formats, must not be "bloated" and the required storage should not exceed by more than 1.5× the raw data format.

Priority: 2

2.3.3 I/O Speed and Efficiency

2–3.3–R1 I/O of data must not be a bottleneck for processing. The total package permformance must follow standards set by the Project (see 4).

Priority: 1

2–3.3–R2 I/O failure (e.g. due to full disks, device offline) should not result in catastrophic failure of the Package. Error recovery should be as robust as possible.

Priority: 1

2–3.3–R3 The Package must be able to handle, efficiently and gracefully, datasets larger than main memory of the host system.

Priority: 1

2.3.4 ALMA Interferometer Data

2–3.4–R1 Correlation products accumulated at multiple bit depths (16-bit,32-bit) or compressed data must be supported transparently (see ALMA SW-11 3.2.3-R5).

Priority: 1

2–3.4–R2 Total power detector data must be handled.

Priority: 2

2–3.4–R3 Phased array data sequences must be handled.

Priority: 2

2.3.5 ALMA Single Dish Data

2–3.5–R1 Data taken with nutating secondary must be supported, as a function of nutator position.

Priority: 1

2–3.5–R2 Total power detector data must be handled.

Priority: 2

2.3.6 Images and Other Data Products

2–3.6–R1 Standard multi-dimensional images must be supported, such as:

2–3.6–R1.1 Spectra and image slices (1D)
Priority: 1

2–3.6–R1.2 Planar images (2D)
Priority: 1

2–3.6–R1.3 Spectral and Time Cubes (3D)
Priority: 1

2–3.6–R1.4 Higher-dimensional Arrays (4D+)
Priority: 2

2–3.6–R2 Other standard derived data products must be supported, such as:

2–3.6–R2.1 Standard Models (e.g. CLEAN models, Gaussian models, disks)
Priority: 1

2–3.6–R2.2 Special Models (e.g. wavelets, Pixons)
Priority: 3

2–3.6–R3 Blanking of pixels must be maintained through the processing of images.

Priority: 1

2.3.7 Foreign Data

2–3.7–R1 Data produced by other interferometers and single dishes in similar observing modes should be importable and processable if provided in a standard data exchange format.

Priority: 1

2–3.7–R2 Imaging data in standard formats (e.g. FITS) from astronomical instruments at different wavelengths should be importable, with the ability to combine (coadd) these with ALMA data where appropriate. This should be through a set of widely used formats, with a **minimal list of supported standards established by the project**.

Priority: 1

2.3.8 Interaction with the Archive

2–3.8–R1 Access from the archive (when such access is granted, e.g. when Package is run by ALMA staff) must be supported.

Priority: 1

2–3.8–R2 The interface between the package and archive must be able to provide data access (when such access is granted) without interfering with other access to the archive.

Priority: 1

2–3.8–R3 Security and integrity of the archive must be ensured during these operations.

Priority: 1

2.4 Calibration and Editing

2.4.1 General Calibration and Editing Requirements

2–4.1–R1 The package must be able reliably handle all designated ALMA standard calibration modes, possibly including but not exclusive to temperature controlled loads, semi-transparent vanes, apex calibration systems, WVR data, noise injection, fast-switching calibration transfer, planetary observations.

2–4.1–R2 Calibration, editing, flagging, and correction of data should be easily reversible within the Package (ie. not requiring re-reading of the data from the archive).

Priority: 1

2–4.1–R2.1 Logging of editing steps should be clearly marked in some sort of history table (possibly distinct from a more readable history).
Priority: 2

2–4.1–R2.2 Individual edit undo is desirable.
Priority: 3

2–4.1–R3 Data editing and flagging should be possible based upon array and environmental monitoring data.

Priority: 1

2–4.1–R4 Data calibration, correction and flagging should be possible based upon standard or user-defined models in either functional or tabular form.

Priority: 1

2–4.1–R5 Calibration should involve flexible averaging of data and calibration quantities with user-controllable interpolation, filtering, weighting, and application scope.

Priority: 1

2–4.1–R6 Interactive data editing, calibration, and display of calibration quantities should be largely graphical and intuitive, with user-definable setups. Specialized editing display tools should include:

2–4.1–R6.1 specification of data by selection of antenna, baseline, time range, uv range, pointing center, slice through data cube, etc.
2–4.1–R6.2 displays of spectra and spectral cubes
2–4.1–R6.3 the above with baseline, time, band and/or channel averaging
2–4.1–R6.4 for interferometer data, amplitude (phase) vs. time on each baseline (Difmap vplot), vs. time-baseline (AIPS TVFLG) with interactive zoom and clipping
2–4.1–R6.5 editing based on difference from a running mean or median, or rms in boxcar, or difference versus model, etc.
2–4.1–R6.6 autoscaling or fixed scaling of colormap, intensity or axes
2–4.1–R6.7 inclusion and marking of flagged data in plots and in autoscaling

Priority: 1

2–4.1–R7 Automatic editing tools should be available in the package, including:

2–4.1–R7.1 editing based on difference from a running mean or median, or rms in boxcar, or difference versus model, etc.
2–4.1–R7.2 outlier editing
2–4.1–R7.3 flagging based on strict selection (e.g. time, baseline, band, channel, etc.)
2–4.1–R7.4 flagging based on ancillary data (e.g. wind speed, sun angle, temperature, etc.)

Priority: 1

2–4.1–R8 Access to time history of calibration information such as source catalogs containing flux density histories, planetary ephemerides, noise tube values, etc. should be built into calibration engines. Output of calibration procedures should be exportable into similar structures.

Priority: 1

2–4.1–R9 Data display and editing should be effected through generic tools applicable to both single-dish and interferometer modes. These should present should, as far as possible, present similar interfaces to the user and have the same look-and-feel.

Priority: 2

2–4.1–R10 Editing should be incorporated into most visualization tools where data or data-derived quantities are plotted, such as from calibration solutions, amplitude vs. uv-distance plots, or any number of other plots. A "see-it, flag-it" capability should be the standard within the tools.

Priority: 2

2.4.2 Interferometer Data

2–4.2–R1 Antenna-based determination of calibration quantities such as gains, polarization leakages, bandpasses, is the primary form of calibration where appropriate.

Priority: 1

2–4.2–R2 In addition to antenna-based calibration, baseline dependent corrections will also be supported. For example, coherence loss due to atmospheric phase fluctuation depends on baseline length (this aspect will be more important at higher frequencies) and must be taken into account if some of the WVR corrections are discarded while others are applied. Also, in general, the bandpasses are baseline dependent and contain non-closing terms.

Priority: 1

2–4.2–R3 Gain corrections will be made based on differences between observed and modeled data quantities, possibly with iteration (e.g. self-calibration and determination of gains using calibration sources). Where solutions are discrepant or poor, automatic editing should be possible.

Priority: 1

2–4.2–R4 Calibration quantities (possibly stored in tables or data structures) should be transferable between sources and/or frequency bands, after any necessary interpolation, extrapolation or smoothing. This will be the primary method of phase calibration transfer using fast-switching between source and calibrator.

Priority: 1

2–4.2–R5 Determination of, correction for, and examination of closure errors should be straightforward to carry out.

Priority: 1

2–4.2–R6 Determination of the complex bandpass using calibration source observations, and transfer to target sources, should be simple and robust.

Priority: 1

2–4.2–R7 Determination of polarization calibration quantities such as leakage (D-term or Jones matrix) and complex gain difference must be an integral part of the package, using both linearized and full matrix calculations.

Priority: 1

2–4.2–R8 Incorporation of standard models (e.g. planetary disks, models for HII region structure, known source spectra) should be easy for calibration operations.

Priority: 1

2–4.2–R9 Redundancy (e.g. same, similar, or crossing baselines) should be used wherever possible to increase accuracy of or check calibration solutions. Editing based on this comparison should be possible.

Priority: 2

2–4.2–R10 Interferometric pointing, focus, baseline, and beam response fitting should be available in the Package as a supplement to the on-line calibration.

Priority: 3

2.4.3 Single Dish Data

2–4.3–R1 Straightforward and flexible fitting of spectral bandpass from calibration source observations is required.

2–4.3–R2 Calibration of temperature controlled loads and noise sources from observations of celestial sources should be supported.

2–4.3–R3 Final data scaling must be possible in case that the 1calibration fails for unexpected reason. A user may want to make their manual scaling, e. g. by referring to the line intensity of the map reference center, to compensate for the daily/time variation.

2–4.3–R4 For single-dish OTF observations with the two orthogonal scan directions (e.g., N-S and E-W), the intensity scale can be adjusted to minimize striping due to slight variation of gain.

2–4.3–R5 Processing for pointing, focus, tipping, or beam-fitting data must be available.

Priority: 3

2.4.4 Mosaicing Considerations

2–4.4–R1 Individual data points must be associated with pointing center information, and one must have the ability to deal with complex scanning strategies.

2–4.4–R2 Determination of and correction for pointing offsets and the beam shape is critical to the ability to reliably mosaic using ALMA, and thus must be available in several algorithmic forms in the package.

2–4.4–R3 The complex polarization response of the telescope beams must be calibratable (though this is mostly an imaging step).

2–4.4–R4 Careful cross calibration of the flux scales between ALMA interferometric data and single dish data is required for high fidelity imaging (important and more difficult for ACA data). There must be tools to cross-check and correct the relative calibration between mosaics and different component observations.

2.4.5 Ancillary and Diagnostic Data

2–4.5–R1 Environmental data such as weather (e.g. wind speed, temperature, dew point) should be available for editing or calibration procedures, and easily incorporated into user-specified calibration models.

2–4.5–R2 Engineering monitoring information such as temperature sensor readings and tilt-meter outputs, perhaps included as ancillary tables attached to data files with special keywords, should be readable and incorporated into the calibration and editing process.

2–4.5–R3 Output from the atmospheric monitoring (e.g. WVR, FTS) instrumentation should be processed and used by calibration software.

2–4.5–R4 Pointing, focus and subreflector information must be dealt with appropriately.

2.5 Imaging

2.5.1 General Imaging Requirements

2–5.1–R1 Imaging data selection from any combination of ALMA exported data, the ALMA archive, or other instruments supporting common export formats must be provided.

2–5.1–R2 Efficient selection of subsets of the imaging data must be provided.

2–5.1–R3 Provision must be made for the utilization and development of a variety of imaging, deconvolution, and analysis algorithms (e.g. flavors of CLEAN, MEM, linear and non-linear mosaics)

2–5.1–R4 Astrometric accuracy must be preserved over phase-calibration distances of at least 5 degrees.

2–5.1–R5 Images made on the different equinox (e.g. B1950 and J2000) or different coordinate (RA,DEC and l,b) system or different projection (tangent, sinusoidal, ...) can be merged and compared appropriately.

2–5.1–R6 Data cubes using different velocity definition (optical or radio definition for Doppler velocity) must be merged appropriately.

2–5.1–R7 Image pixel blanking should be supported.

2.5.2 Interferometer Imaging

2–5.2–R1 High-fidelity imaging of the entire primary beam in all Stokes parameters is the primary goal — therefore, incorporation of the polarized primary beam response of the array is required.

2–5.2–R2 Imaging must deal seamlessly with mosaiced data, with proper gridding in the uv-plane and compensation for primary beam effects and pointing in such a manner as to mitigate the effects of non-coplanar baselines and sky curvature. A variety of options for gridding and beam correction should be available at user request.

2–5.2–R3 There must be seamless integration of data from multiple epochs and configurations

2–5.2–R4 There must be the ability to include short-spacing data taken in single-dish mode (both ALMA and non-ALMA data)

2–5.2–R5 Subtraction of continuum level from spectral data is required. This can be done in both the Fourier and image domain. In the case of uv-plane subtraction, flexible setting of the frequency channel ranges for the calculation of the continuum level (graphically as well as CLI) should be available.

2–5.2–R6 The creation of 3D images for rotating object (e.g. planets) should be supported.

2.5.3 Single Dish Imaging

2–5.3–R1 The package must be able to produce an image by using data observed at different spacing (or even at random positions); must be spatially interpolated or re-gridded correctly (e.g. 1 arcmin RA-DEC grid observations and 1 arcmin l-b grid observations; two sets of 1 arcmin grid observations using two different map reference centers (0,0) and (0'.2, 0'.6))

2–5.3–R2 The package should have the capability to produce the stamp map (profile map) of spectra at a grid of positions.

2.5.4 Mosaicing Considerations

2–5.4–R1 Combination of interferometer and single-dish data into mosaic imaging is essential.

2–5.4–R2 Careful (polarized) primary beam correction and pointing correction is critical for high fidelity mosaic imaging and must be incorporated into the mosaicing algorithms.

2.5.5 ACA Imaging Considerations

2.6 Analysis

2.6.1 General Analysis Requirements

2–6.1–R1 The astronomer must have the capability to develop their own tools or tasks, with easy access to data and images, and straightforward interface with the package

2–6.1–R2 Translation between various astronomical quantities and units (e.g. Jy and K, MHz and km/s) should be straightforward and user selectable.

2.6.2 Visibility Data Analysis

2–6.2–R1 UV-plane analysis based on goodness-of-fit to a model will be required.

2.6.3 Single Dish Data Analysis

2–6.3–R1 Automatic measurement of line parameters (line intensity, integrated intensity, Gaussian-fit line width, rms noise level, ...) for user specified velocity (frequency) window must be made, and to be stored in a text-format list file that can be output by the user if desired.

2–6.3–R2 Effective, robust and precise spectral baseline removal facility is required. Fourier analysis of standing waves and their removal will be a critical task.

2–6.3–R3 Fourier transform to a pseudo-uv plane image will be needed.

2.6.4 Mosaicing and Combined Array Analysis

2–6.4–R1 Seamless transformation between image-plane and uv-plane analysis is necessary.

2.6.5 Image Analysis and Manipulation

2–6.5–R1 The ability to extract lower-dimensional slices from n-dimensional data cubes efficiently is required.

2–6.5–R2 The ability to collapse or integrate over sub-dimensions of data cubes in order to form "moments" is required.

2–6.5–R3 Blanking of pixels (magic-value) must be maintained through the analysis process. It is desirable that blanks not be destructive (the original pixel value is retained), and it be possible to turn on and off different blanking (mask) levels.

2–6.5–R3.1 Interactive and automatic facilities for setting of blanking parameters (e.g. windowing, S/N based blanking) to avoid degrading S/N in the analysis must be provided.

2.7 Visualization

2.7.1 General Visualization and Plotting Requirements

2–7.1–R1 Plotting and display capabilities should be integrated into the GUI tools throughout the package. Where possible the displays should have similar look and feel to reduce the plotting learning curve.

2–7.1–R2 Plotting of data and calibration quantities as a function of time (LST,UT,etc.) in both standard X-Y and image cube (with additional axes) must be available.

2–7.1–R3 Where appropriate, edition and flagging capability should be incorporated into all plots. Basically, if you "see-it" you should be able to "flag-it".

2–7.1–R4 The output of the display should be possible in many different designated formats, including fits, postscript, pdf, gif and jpeg.

2.7.2 Display Appearance and Interactivity

2–7.2–R1 User should be able to produce overlays of different data sets of standard formats. It should be possible to place these data sets in layers which can be switched on and off separately. The different images should be editable, and it should be possible to declare certain colors transparent. It must be possible to shift, rotate and scale the images at will.

2–7.2–R2 It must be possible to display and overlay data with different coordinate systems, i.e. the coordinate system of the display can be chosen independent of the system the data were observed in.

2–7.2–R3 Both contour plots with variously colored and styled lines and false color maps should be possible, it should also be possible to produce RGB overlays (i.e. one layer gets assigned intensity scales of red, another one of green, and one of blue), or Hue/Saturation/Value.

2–7.2–R4 User should be able to manipulate intensity and color scales easily and graphically, the setup achieved should be saveable and reloadable.

2–7.2–R5 User should be able to add annotation, both interactively and through scripts, for publication quality plots, i.e. text with various fonts (including Greek letters), symbols (e.g. all the symbols provided by the LaTeX package with AMSTeX extension), arrows, geometrical figures like boxes and circles etc. Different line styles, sizes, thicknesses and colors for all those should be available. The various elements should be editable and removable separately, and it should be possible to put them in a separate layer.

2–7.2–R6 The display tool should have astrometry facilities, i.e. based on catalogs and by assigning sources in the maps it should be possible to calculate the coordinate system.

2–7.2–R7 The display for spectra must be linked to molecular data bases which make identification of the lines possible.

2.7.3 Image-cube Manipulation

2–7.3–R1 Interactive display of spectra corresponding to a displayed image should be supported. For example, the display of a spectrum by clicking the the image map on the display, for the position nearest the cursor position. Also, dragging the line on the map to bring up a position-velocity diagram.

2–7.3–R2 Plotting of spectra on a pseudo-grid corresponding to position on a raster (e.g. a "stamp map" or "profile map", basically thumbnail spectra in panels corresponding to position) should be possible.

2–7.3–R3 Data cubes should be viewable as movies with selectable frame rates.

2–7.3–R4 It should be possible to view arbitrary subsets or slices of data cubes.

2.8 Special Features

2.8.1 Simulation

Note: Inclusion of at least moderate simulation capability in the Package will benefit the user who may not have access to the more comprehensive simulator built into the Pipeline and Online systems. In the spirit of providing the user with offline options equivalent to those available in the project, it is desirable that some simulation capability also makes it into the Package albeit at low priority. The main goals would be to allow the user to simulate basic ALMA modes based on input models or images and to replace existing data in ALMA format with simulated data.

2–8.1–R1 The Package shall inherit the basic simulation capability of the ALMA Online Data Processing Pipeline (Levels 1 and 2) as outlined in ALMA Memo SW-11, as per (5).

Priority: 2

2.8.2 VLBI

Note: It is assumed that VLBI data will be reduced in whatever package supports general VLBI observations, as ALMA will not be a stand-alone VLBI instrument. There are no VLBI requirements on the ALMA Offline Data Processing Package, though it may end up being the case that the Package also supports VLBI processing. This would have some advantage to users of the ALMA Packgage, and thus it is worth including a low-priority requirement to this effect.

2–8.2–R1 It is desirable that the Package also support general VLBI processing suitable for dealing with ALMA data, in order to allow general ALMA users access to VLBI science without having to learn a completely new software package.

Priority: 3

2.8.3 Solar Observing

Need author for this section

2.8.4 Pulsar Observing

Need author for this section