



Title:	Author:	Date:
NRAO Doc. #:		Version:

SRDP/SRAO Wave 3 Conceptual Requirements

PREPARED BY	ORGANIZATION	DATE
John Tobin	NRAO	

Change Record

VERSION	DATE	REASON
1.1	01/08/2021	Additions and reorganization for improved clarity

Table of Contents

1 Introduction.....	3
Purpose.....	3
Specific SRDP/SRAO Project Goals.....	3
Scope.....	4
References.....	4
Abbreviations and Acronyms.....	4
2 Requirements To Pipeline and CASA.....	4
VLA Calibration Improvement.....	4
Wave 2 Deliveries from Pipeline.....	5
Carryover Priorities.....	5
New Wave 3 Priorities.....	5
Stretch Goals.....	5
VLA Spectral Line Calibration.....	5
Wave 2 Deliveries from Pipeline.....	5
Carryover Priorities.....	5
New Wave 3 Priorities.....	5
Stretch Goals.....	6
VLA Imaging Pipeline.....	6
Wave 2 Deliveries from Pipeline.....	6
Carryover Priorities.....	6
New Wave 3 Priorities.....	6
Stretch Goals.....	6
ALMA User-Defined Imaging.....	7
Wave 2 Deliveries from Pipeline.....	7
Carryover Priorities.....	7
New Wave 3 Priorities.....	7
Stretch Goals.....	7
3 Requirements to the SSA.....	7
3.1.1 ALMA image products.....	7
.....	7
Carryover priorities.....	7
3.1.2 Images Display in Archive.....	8
New Wave 3 Priorities.....	8
3.1.3 CARTA Integration.....	8
Priorities in Development.....	8
Carryover Priorities.....	8
Progress Report.....	8
3.1.4 User-Defined imaging Interface Improvements.....	8
Carryover Priorities.....	9
New Wave 3 Priorities.....	9
3.1.5 Large Project Support.....	9
Carryover Priorities.....	9
Progress Report.....	9
3.1.6 Catalog Display and Search.....	9

Carryover Priorities.....	9
Progress Report.....	9
3.1.7 Improved Archive Metadata.....	10
Carryover Priorities.....	10
New Priorities.....	10
Progress Report.....	10
3.1.8 Image cutouts.....	10
Carryover Priorities.....	10
Progress Report.....	10
3.1.9 Data Delivery.....	10
Carryover Priorities.....	10
New Wave 3 Priorities.....	11
3.1.10 Cache system.....	11
Carryover Priorities.....	11
Progress Report.....	11
3.1.11 Observatory Database.....	11
Carryover Priorities.....	11
Progress Report.....	11
3.1.11 Astroquery Support for Archive.....	12
New Priorities.....	12
3.1.12 New Archive Frontend.....	12
New Priorities.....	12
Progress Report.....	12

1 INTRODUCTION

Purpose

This is a Science Ready Data Products (SRDP) Project Office document describing the conceptual requirements for the wave 3 capabilities of the SRAO (Science Ready Archive and Operations) project of the SRDP Program. SRAO releases functionality in a rolling wave approach with major features released on an approximately yearly basis. SRDP is currently in the Wave 1 phase, including VLA SRDP Operations, calibrated MS delivery (VLA and ALMA), and ALMA User-Defined Imaging (AUDI). With the launch of SRDP operations for Wave 1 and Wave 2 development coming to a close, it is time to plan the Wave 3 requirements to be developed over the course of 2021.

Specific SRDP/SRAO Project Goals

- VLA Calibration
 - Improve the standard calibration of VLA continuum data to detect and flag more instrument problems during execution to reduce flagging required by DAs
 - Improve heuristics for standard VLA calibration to result in improved outcomes from the calibration pipeline
 - Enable VLA continuum and spectral line calibration as part of a single workflow
 - Bring polarimetric calibration into PI pipeline
- SRDP Imaging via NRAO archive
 - Provide a simplified interface for re-imaging ALMA and VLA data that fulfills the needs of many general users
 - Provide some adjustable image parameters (resolution, channels, and cube width)
 - Enable automatic self-calibration to improve image quality for general users
 - Define an imaging pipeline for the VLA to produce science-ready data, on par with the ALMA imaging pipeline
 - Improve overall speed of the system by aggregating multiple requests
 -
- Facilitate data discovery in NRAO archive
 - Improve metadata display for ALMA and VLA data
 - Enable online viewing of data with CARTA browser
 - Enable users to explore standard data products available from ALMA archive
 - Serve data from Large VLA, ALMA, and VLBA projects; large projects from other observatories may also be considered
 - Enable scripted access to archive holdings
 - Enable users to request recalibration of archival ALMA and VLA data with the new pipelines

Scope

This document details the conceptual requirements for the third wave of major SRDP functionality. The conceptual requirements are sub-divided into functionality that is primarily related to CASA and ALMA/VLA pipeline functionality and requirements that are primarily related archive or system level functionality. Interdependencies will be pointed out, for example, it may be needed for the archive interface to pass a parameter to the pipeline in order to activate/deactivate functionality. All items listed are high priority for the project and the prioritization of implementation will be coordinated with DMS, CASA, SSA, and the Pipeline Team. A comprehensive summary of bugs discovered and/or fixed in existing software or minor tweaks to existing functions are not included in this document.

References

[SRDP System concept]

[SRDP System Requirements]

Abbreviations and Acronyms

ALMA Atacama Large Millimeter/submillimeter Array

CASA Common Astronomy Software Applications

QA Quality Assurance

RFI Radio Frequency Interference

RMS root-mean-squared noise

SB Scheduling Block

SPW Spectral Window

SRDP Science Ready Data Products

SSA Science Support and Archive

VLA Karl G. Jansky Very Large Array

VLA Sky Survey

2 REQUIREMENTS TO PIPELINE AND CASA

The SRDP requirements from the pipeline are divided into four main categories, VLA calibration, VLA spectral line calibration, VLA Imaging, and ALMA Imaging.

VLA Calibration Improvement

During Wave 2, we were able to achieve some of our highest priorities like enabling multi-band calibration for the pipeline. However the infrastructure work was more extensive than initially estimated and other development items could not be scheduled as a result. As such, we highlight what was done in Wave 2, carryover priorities from Wave 2 and new priorities. We also identify some stretch goals that

may have some work done toward requirements definition/implementation during Wave 3 and could be Wave 4 priorities.

Wave 2 Deliveries from Pipeline

- Multi-band calibration
- Bad deformatter on reference antenna recovery
- Multiple bandpass calibrator fix

Highest Priorities

- Completion of AQUA reports for VLA/VLASS*
- Enhanced DTS issue identification and flagging (CASR-563)*
- Improved calibrator RFI flagging heuristics (PIPEREQ-12; PIPE-939)*
- Gain table flagging extensions
 - Ripple detection (bad switch problem) (PIPEREQ-11)
 - Gain outlier detection (PIPEREQ-10)
- Target RFI Flagging (CASR-540)
- Per-baseband delay solutions (PIPE-670)*
- Spectral window mapping for gain calibration (PIPE-671)*
 - enables phase calibration if not enough S/N due to faint calibrator*
- Multi-band weblog updates (PIPE-935)
- Detect high delays (and scatter) on reference antennas (PIPE-1032)

Medium Priorities

- Allow different bandpass calibrators per band (PIPE-669)
- Antenna position corrections in a post-observation time window (CASR-556)
- Fluxboot gain solution flagging (PIPEREQ-9)
- Fluxboot QA metric adjustments (PIPEREQ-8)
- Enable compression correction for S-band in PI pipeline

Low Priorities/If Time available

- Generalize the hifv_cirfeedpolcal pipeline task for all bands (polarization calibration)
- Fix pipeline bugs to allow for experimental P-band calibration with the pipeline

* - denotes carryover requirement from last wave planning

VLA Spectral Line Calibration

Little progress was made toward spectral line calibration during Wave 2 implementation, but progress has been made in defining requirements. The VLA pipeline currently has a number of shortcomings with regard to spectral line calibration that require manual intervention. The primary goal is to develop a spectral line reduction recipe that can be run on datasets by PIs, as a recalibration request, or by the DAs for SRDP processing.

Wave 2 Deliveries from Pipeline

- Bugfix for cont.dat interpretation (correctly transforms LSRK to TOPO)

Highest Priorities

- RFI flagging on uv-grid (CASR-496)
 - avoids flagging of bright spectral lines
- Enable data weights to be determined from switched power (CASR-476)*
- Design and set up independent spectral line calibration recipe that could be used by end users, triggered using recalibration workflow, or by SRDP DAs
- Per-baseband delay solutions (PIPE-670) – cross-list with standard calibration*

Medium Priorities

- Use MS tool hanning smoothing to enable per spw hanning smoothing (CASR-523; PIPE-672)*
- Spectral window mapping for gain calibration (PIPE-671) – cross-list with standard calibration*
 - enables phase calibration if not enough S/N on narrow spws*
- Cont.dat interpretation to treat non-specified spws as continuum-only

Low Priorities/If Time Available

- Enable pipeline to use multiple MS files to perform continuum processing/imaging on one
 - ALMA has similar requirement for this cycle
- Carry out spectral line optimized calibration on the other when spectral lines are a science goal set in PST – relies on a connection of OPT database to processing environment via SSA

* - denotes carryover requirement from last wave planning

It would be preferable to know if a dataset is meant to be used for spectroscopy and if so, what spectral windows are intended for spectroscopy. In the absence of this information, which needs to be passed from the PST into the metadata of the pipeline environment, we need the ability to process the entire dataset under the assumption that it is spectral line, while also creating a continuum-optimized reduction. Within this effort we will be exploring the possibility of using the switched power measurements to set the weights for the VLA data, which may be useful for both calibration and spectral line imaging.

VLA Imaging Pipeline

The work on a basic continuum imaging pipeline has been completed and continuum images can now be produced by the pipeline with proper imaging heuristics. However, the current pipeline has shortcomings that need to be improved upon. Automasking is not currently used (it was descoped and we bring it back in as a new priority) and cleaning is done without a mask to a high-threshold (5-sigma). Furthermore, work was not able to be done to detect mosaicked fields for the VLA. Finally, we need to enable a basic cube imaging pipeline.

Wave 2 Deliveries from Pipeline

- Basic continuum imaging pipeline (CASR-543,PIPE-673)

Highest Priorities

- Evaluate automasking heuristics for VLA continuum imaging pipeline

- Main work will be in determining the optimal settings for the automasking
- Detect mosaicked fields and perform joint deconvolution with mosaic gridded*

Low Priorities/If Time Available

- Define requirements for the continuum imaging pipeline to use outlier fields to remove effects of bright sources rather than using large images
- Develop heuristics for using wproject gridded for single pointing S and L-band images
- Enable basic cube imaging

ALMA User-Defined Imaging

The ALMA User-Defined Imaging recipe is functioning well and we are now focused on adding additional features to the service. The primary new feature is user-specified angular resolution, the pipeline is now capable of this using the hifas_imageprecheck task, and work to the archive interface needs to be done to allow for parameter input and pass it to the pipeline.

Wave 2 Deliveries from Pipeline

- User specified angular resolution, with uv-tapering enabled (PIPE-708) – requires archive interface to enable use of feature by users

Highest Priorities

- Line + continuum imaging (CASR-509)*
- Characterize taperability of ALMA data to validate limits of uv-tapering applied.
- Define requirements to enable self-calibration*
- Image QA using continuum-subtracted MFS images*

Medium Priorities

- Enable experimental self-calibration of ALMA datasets.

Low Priorities/If Time Available

- Define requirements to enable multi-configuration imaging with ALMA (12m-only)

3 REQUIREMENTS TO THE SSA

The NRAO archive, developed and maintained by the SSA team constitutes the user-facing portion of the SRAO project. Requirements provided here are to facilitate user interactions with archive functionality (AUDI imaging) and enable data discovery. Items beyond 3.1.9 are listed for completeness, but are lower priority than 3.1.1 through 3.1.8.

We provide Carryover priorities for work that has not been finished for Wave 2, New priorities, and a Progress report (if there is progress to report on a specific goal). We list the overall prioritization first and the requirements for each priority are described in more detail in the following sections.

Highest Priorities

- CARTA Integration
- ALMA Image Product Ingestion
- Image Product Display Improvements
- AUDI Interface Enhancements
- Large Project Support
- Catalog Support

Medium Priorities

- Data delivery enhancement
- Astroquery support
- Observatory Database

Low Priorities

- New Archive Frontend
- Image cutout service
- Cache System

3.1.1 ALMA image products

To facilitate data discovery, the ALMA image products need to be ingested and served from the NRAO archive.

- All standard ALMA imaging pipeline products should be ingested into NRAO archive
 - Standard ALMA products should be identified relative to SRDP products
 - Calibrator images should be ancillary
 - per spw MFS images should be ancillary
 - Thumbnails should be provided for images
 - harvested from weblogs if possible, aggregate continuum and mom8 for lines
 - Images should be viewable in CARTA from the archive without download
 - Ingest Manual images if possible and if deemed suitable

3.1.2 Images Display in Archive

Image in the current archive are not presented on equal footing as visibility data and there are a number of peculiarities with the way images are presented to the users in the archive interface that should be refined.

- Implement suggested fixes from (SSA-6453)
 - Display images as search for by users

- Display correct and additional metadata needed to describe images
- Image search should return points within an image field of view, not just relative to an image center

3.1.3 CARTA Integration

The ability to explore ALMA,VLA, and VLASS data online, without downloading extremely large datacubes is essential for the user experience. The client-server architecture of CARTA naturally lends itself to this use case.

Requirements in Development

- The users should be able to select an image in the archive interface and have it open in a CARTA session in their web browser while taking comparable time to CARTA in the ALMA Science Archive.
 - This is the most important requirement

Additional Requirements

- The user shall be able to select a second image and have it open/overlaid in the same CARTA session.
- The user shall be able to select a region in the CARTA session and use this information in the archive interface (either for cutouts or as a search parameter).

Progress Report

The SSA team has developed a prototype CARTA integration and demonstrated it. The next steps are to prepare the prototype for production.

3.1.4 User-Defined imaging Interface Improvements

The current interface for requesting images from ALMA data is rather simple and some improvements can be made to better serve the users by helping them find the information they need to make new cubes. We also need an interface developed for VLA imaging, derived from the ALMA imaging, as well as revision to support new features provided by the pipeline.

- Enable user-specified angular resolution (highest priority, pipeline development finished; SSA-6689)
 - Add dialog box for user to specify angular
 - Pass user-specified resolution to template recipe
- Allow selection of multiple sources and/or spectral windows for making cubes
 - Already supported by pipeline recipe
- Create interface for VLA User-Defined Imaging
 - Continuum-only at present, simplified interface
- Additional imaging options and modes(selectable through user interface)
 - Cube imaging without continuum (high priority, available in current recipe)
 - Continuum imaging only (high priority, needs another recipe)
 - continuum+line imaging (medium priority, needs recipe and pipeline development)
- Useability improvements

- button to either open Splatalogue in a new tab for the frequency range of the current SPW or an in-line display of molecular lines in the given spectral range from Splatalogue. (medium priority)
- Ability to select from a dropdown list of commonly used rest frequencies within the given SPW, populating this from something like the 'Quick Picker' molecules on Splatalogue. (medium priority)
- Require ability to only allow SRDP QA'd datasets to be imaged
- <https://open-confluence.nrao.edu/display/SRDP/VLA+User+Driven+Imaging+Requirements>

3.1.5 Large Project Support

- Create mechanism for Large VLA, VLBA, and ALMA projects to upload their data products to the archive.
 - Project-specific metadata should be possible to define.

Progress Report

Progress has been made for the support of Large projects as specific 'Collections' in the archive. VLASS and RealFast are the prototypes of this capability. The ingestion mechanism for external projects needs to be developed.

3.1.6 Catalog Display and Search

- Catalogs need to be ingested into the archive and should be able to be searched upon. Catalogs should also be able to be overlaid within CARTA seamlessly.

Progress Report

This capability is expected to be brought online alongside Large Project Support.

3.1.7 Improved Archive Metadata

- **QA state**
 - The archive must display the QA state for the data products contained within the archive, encompassing the differing QA systems for ALMA, SRDP imaging on ALMA data, and SRDP products for the VLA.
- **Weblogs**
 - The weblog from product calibrations for both ALMA and VLA calibration and imaging should be available for immediate viewing from a link.
- **Proprietary state**
 - The archive needs to better track if a dataset is proprietary or not to reduce confusion on which data are available for restoration/imaging.

- VLA Metadata Display updates
 - Necessary for data discovery and imaging use-cases
 - <https://open-confluence.nrao.edu/display/SRDP/VLA+Metadata+Display+Updates>

Progress Report

Submitted to SSA for planning.

3.1.8 Image cutouts

- Provide an image cutout service for archived ALMA, VLASS, and VLA images
 - enable sub-regions to be specified for single images or multiple images
 - provide batch service to upload catalog and return cutouts available in the archive

Progress Report

Other priorities have taken precedence and this feature is not highly useful until the image archive is populated.

3.1.9 Data Delivery

- Enable enhanced data selection and manipulation prior to downloading. XML template recipes exist in the CASA pipelines to support this.
 - Download particular SPWs
- Refactor delivery data structures to be more intelligible
 - <https://open-confluence.nrao.edu/display/SPR/Delivery+Directory+Improvements>
 - Select sources
 - Select intents
 - Time averaging
 - Spectral averaging

3.1.10 Cache system

- Implement a cache system for calibrated measurement sets for more rapid access to data
 - Enables users to re-image data more than once or for different targets without needing a full restore
 - Reduces latency period between image requests and completion of jobs

Progress Report

This has been a low priority item as other priorities have taken precedence.

3.1.11 Observatory Database

An observatory database will be required to fulfill the end goal of SRDP from the proposal process to data processing. The need for such a dynamic database for calibrator flux densities and models is being made more critical by the flux variability of two (3C138 and 3C48) out of the 4 primary flux density calibrators for the VLA.

- Calibrator flux densities
 - Dynamically updated
 - accessed by the pipeline
- Valid UV ranges
- calibrator models
 - polarization
 - flux density
 - time-dependent models
 - individual calibrators
 - accessed by the pipeline
- Baseline solutions

Progress Report

Still in requirements capture phase, other priorities have taken precedence.

3.1.11 Astroquery Support for Archive

- Archive interface should support astroquery interface from Python (high priority)
 - enables archive to be accessed from standard Astronomy libraries and will enable easier use of NRAO data products by researchers

3.1.12 New Archive Frontend

- Prepare new archive frontend interface to better facilitate data discovery

Progress Report

Prototype has been constructed, some technical hurdles remain in creating a search return that has a combination of MOUSes, VLA EBs, and Images.