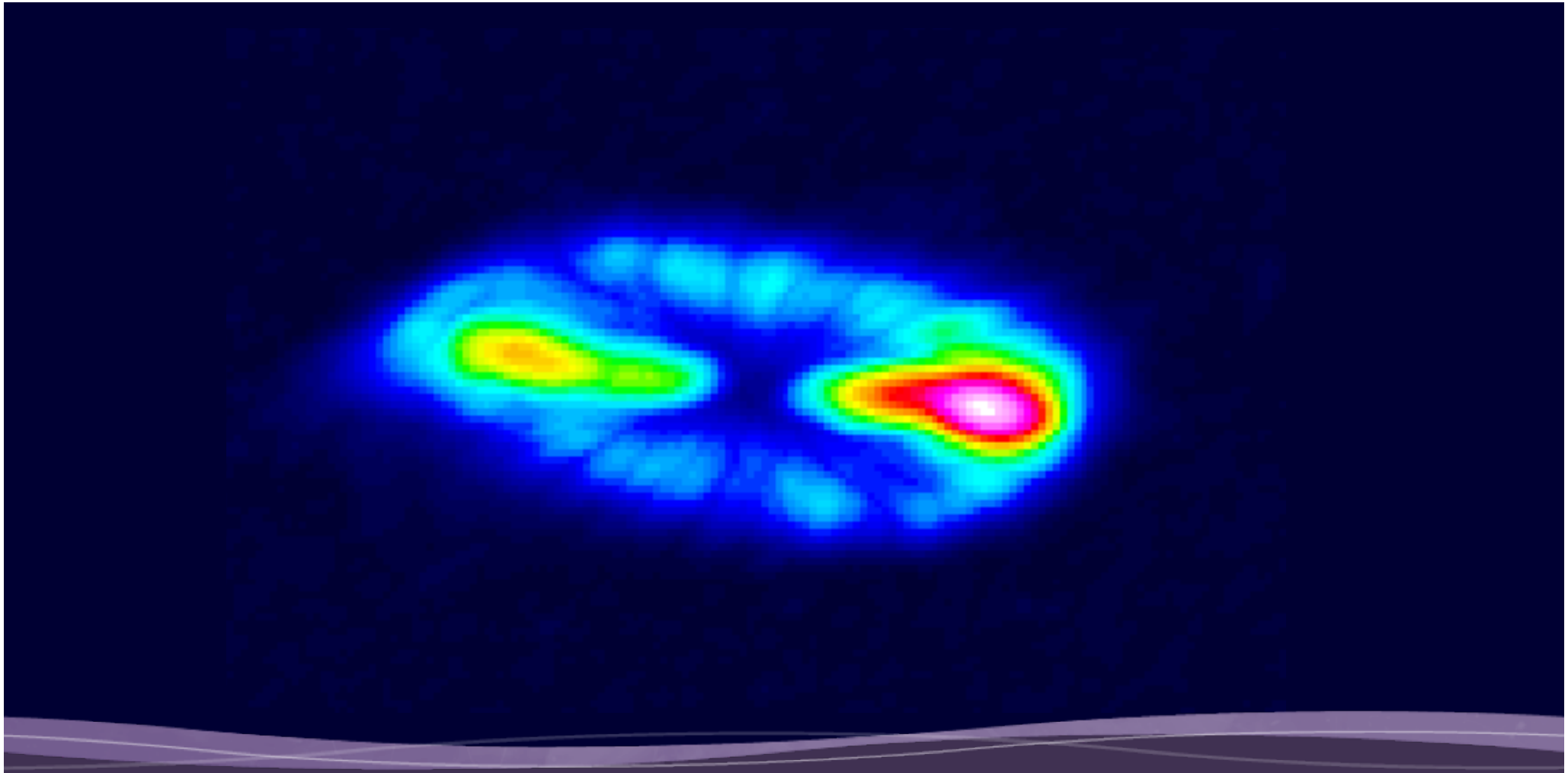


SRDP System Concept

Jeff Kern

NRAO NSF Associated Universities, Inc.

Requirements Process

- System Concept Document
  - Produced by Stakeholder Requirements Committee
  - Review and comment period from NRAO staff
  - Review and comment by UC and CUC

- System concept sets the top level scope of the project.
  - Not detailed enough to actually implement
  - Intended as a 5-year implementation period
    - Will deploy capabilities as they are ready
  - Rolling wave requirements definition

SRDP Requirements Committee

Current Membership

SSR

Eric Murphy

NM OPS

Josh Marvil

Claire Chandler

DMS

Rafael Hiriart

Daniel Lyons

ALMA

Catarina Ubach

Amanda Kepley

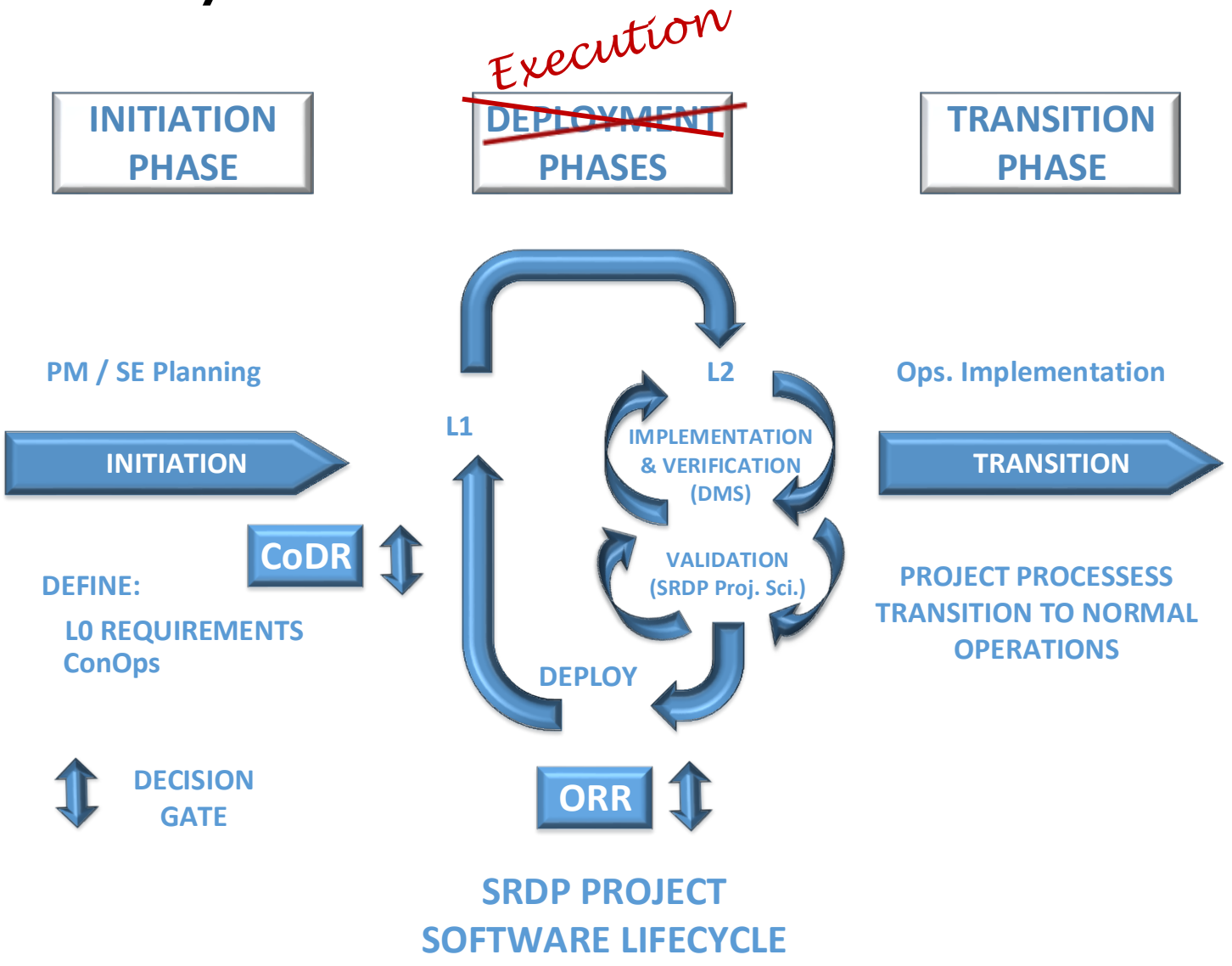
Observers

Walter Brisken

James Robnett

Acting Chair: Jeff Kern

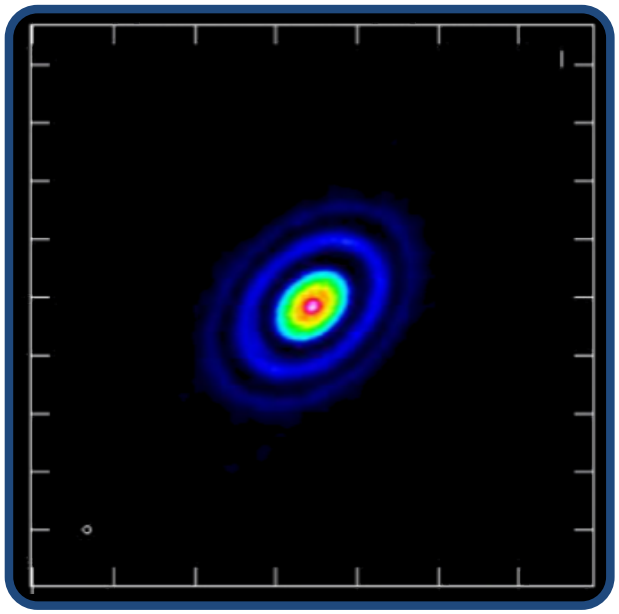
Project Lifecycle

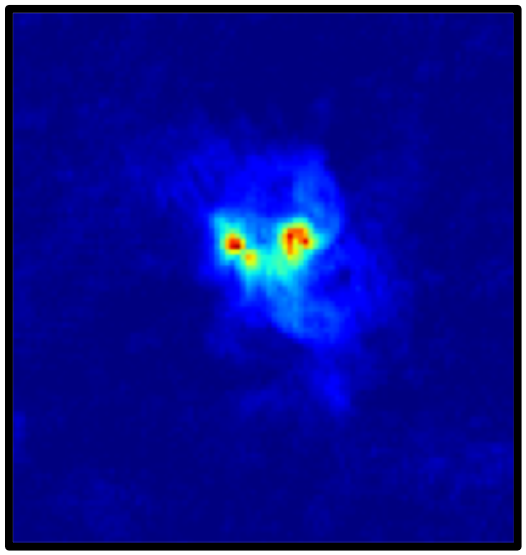


Science Ready Data Products

Key Science Drivers

- The goal of the SRDP project is to increase the scientific impact of NRAO's telescopes by
  - Allowing users to focus more on science and less on data reduction
  - Broaden our user community by decreasing the barriers to using NRAO's interferometers
  - Curating a rich archive of images for archival study





Actors: Functional View

- **Telescope Users**
  - Enter system through observing proposal (PST / OT)
  - Defined objective known products
- **Archive Users:**
  - Enter system through Archive interface.
  - Seek to re-use existing data
  - Require to register with **myNRAO** if they request processing
- **Large Projects:**
  - Subclass of Telescope and Archive Users
  - Significant investments from PI and Observatory
- **Operations Staff:**
  - NRAO staff with effort allocated to supporting the operations of the SRDP processes.

Actors

Experience Level

- **General Public:**
  - Amateur Astronomers, Educators,
  - Primarily interested in the produced images,
    - available for anonymous download.
  - Requirement: Support for file formats they are familiar and prepared to work with.
- **Casual Users / Novice Astronomers**
  - Quickly get an image w/o having to know much about radio astronomy.
    - Download a pre-existing image
    - Easily produce an image from previously obtained data
    - Simple path to obtaining new data and images

Actors

Experience Level (2)

- **Future Power Users:**
  - Start at a basic level, then increase in sophistication
    - Grad students, astronomers w/o radio experience
    - Start as the previous category but over time will interact more deeply with the data.
  - **Requirement**: SRDP interfaces support a continuum of expertise, not just novice and expert modes.
- **Experienced Radio Astronomers:**
  - Current expert users of the NRAO telescopes.
  - Decrease data processing effort to produce timely results
  - Engage these experts to enefit from their expertise in reduction heuristics and quality assessment.
  - Gatekeepers, must accept SRDP products before the wider community will

Actors

Workflow System

- Automated "state engine" that tracks the various workflows.
- Concept is similar to:
  - Survey Status Data Base from VLA Sky Survey
  - ALMA state system (although likely to be much simpler)
- Design of system is TBD by DMS

External Interfaces

- **Proposal Submission and Observation Planning (PST/OPT/OT)**
  - PST/OPT: New requirements to support intent capture, and support for observatory standard observing strategies.
  - ALMA OT: Contains much of the relevant information already
- **Archive Interface**
  - Primary interface for finding, creating and accessing science-ready products.
  - data and product discovery capabilities
  - product inspection facilities
  - interface through which custom products may requested (PPI)
  - and product delivery mechanisms.
- **Weblog**
  - Primary record of the process that generated a set of science products.
  - Interface for quality assurance and detailed inspection by experienced users.

External Interfaces (2)

- **Helpdesk**
  - User / operations staff communication mechanism
  - Paradigm change so that tickets are created by NRAO to initiate communication
  - Customization and integration
    - Prevent need to cut and paste from the AAT or elsewhere
    - Automated update from the workflow engine
- **Workflow Management Interface**
  - Operations Staff interface to monitor, control, and respond to jobs in the SRDP workflow.
  - Combination of Project Tracker and AQUA interfaces in ALMA

Policy Choices

- Calibration Strategy
  - *Not all projects are suitable for SRDP calibration and imaging. Users should be able to "opt out" of using a standard calibration when warranted.*
  - *Observatory-derived SRDPs will only be generated for those projects that conform to observatory "template" observing strategies*
- Archive Contents
  - *All products in the archive shall have full provenance information and have passed a quality assurance process.*
  - *Ingestion of user-generated products is outside the scope of the SRDP project.**
- Proprietary period for products same as the underlying data.
- Do not require all observations to be in full polarization.
- VLA / VLBA stay with guaranteed time observing model.

**Large or commensal projects are excepted from this rule*

Quality Assurance

- ALMA has established QA process for standard calibration and imaging.

- VLA calibration QA will be similar to ALMA
  - RFI will make it more challenging
  - Need additional resources to do ALMA like QA and improve the calibration pipeline.

- Imaging QA is different from ALMA QA2:
  - Not a question of if the observations are suitable.
  - Is the product reliable and well characterized?

- Will work to automate QA as much as possible, but expect a human will look at every product before it goes into the archive.
  - *Unknown unknowns*

SRDP Use Cases

- **Standard Calibration** – automated science-quality assured calibration for supported observational modes.
- **Standard Imaging** – automated science-quality assured images for supported observational modes. Objective is a homogeneous set of images in the archive to support telescope and archive users.
- **Optimized Imaging** – targeted generation of science quality assured images as requested by a science user. Images will be quality assured (in conjunction with the user) and delivered to both requestor and archive.
- **Archive Use** – processes by which science-ready data products are made accessible to user. Includes data discovery, creation of processing requests, and the delivery process.
- **Restore** – returns measurement sets to the calibrated and flagged state they were at the end of the standard calibration process, for delivery to the user or to serve as the initial state for other processes.
- **Recalibration** – the process of redoing a calibration, either with a different version of the supporting software tools, or with additional inputs from the user. Quality assured recalibration products are stored in the archive.
- **Combined Imaging** – takes data taken from multiple configurations (including the ALMA Total Power Array) to produce quality assured images with an increased range of emission scales.
- **Time Critical Observation's** – modifies the standard processes to support triggered and target of opportunity observations.
- **Large Projects** – focuses on maximizing the return to the project team, the Observatory and archive users, and leveraging the special relationship between these projects and the Observatory.
- **Curation and Reproducibility** – describes how the SRDP project will assist the NRAO user community in documenting reproducible data reduction processes.
- **Commissioning and Validation** – covers the special considerations that are necessary for the SRDP staff to commission and validate the SRDP deliverables prior to their release.

Use Case: Standard Calibration

automated science-quality assured calibration for supported observational modes

Does not apply to ALMA!

- For most projects automatically create SBs with observing strategy that conforms to SRDP calibration requirements.
- Pipeline automatically executes at completion of each execution.
- Quality Assurance process augmented to classify and correct errors
  - Most calibration corrections are expected to be additional flagging
  - Necessary to improve the pipeline success at producing SR calibration
- Calibration products made available to the user and stored in the archive

Use Case: Standard Imaging

automated science-quality assured images for supported observational modes.

Does not apply to ALMA!

- Objective is a homogeneous set of images in the archive to support telescope and archive users.
  - VLA process will be developed after the optimized imaging
- Automatically triggered when the observation is complete
  - Need additional project structures to be able to determine this for VLA
- QA and if necessary corrective action by SRDP operations staff
- Products:
  - Full bandwidth Stokes I continuum per band
  - Spectral index (if $\Delta\nu/\nu >$ Threshold)
  - Cubes at PI requested spectral resolution for spectral projects‡
  - Stokes Q and U images for polarization products

‡ *Subject to reasonable size and compute restrictions.*

Use Case: Optimized Imaging

targeted generation of science quality assured images as requested by a science user

- User initiated through Archive interface
- Starting point is Calibrated MS
- Image properties specified through archive interface
  - field of view, spectral extent, spectral and spatial resolution, and polarization
- Image created using imaging pipeline, with manual intervention when necessary.
- Helpdesk ticket created and used to track evolution of image and work with user to validate image
  - *Resource constraints will need to be in place, both staff and computing limitations need to be respected!*
- Images from calibrations stored in the archive will be archived for other users.

Use Case: Archive Interface

processes by which science-ready data products are made accessible to user.

- Primary User Interface to SRDP
  - Must be logged in aside from most simplistic data search and download
- Data Discovery
  - Identify interesting products in the archive
  - Show available related products
    - Available calibrations
    - Images from a data set (or vice versa)
- Product Visualization
  - Thumbnails of images
  - Summary of observations
  - Web based Image Viewer (CARTA)

Use Case: Archive Interface

processes by which science-ready data products are made accessible to user.

- Data Selection
  - "Shopping Cart" for multiple products
  - Download or select for further processing

- Data Delivery
  - Select method of data delivery
    - Download manager, stage to Lustre, disk shipping(?), stage to S3(?)
  - Some processing available prior to download
    - Selection, averaging, moments(?)

Use Case: Restoration

return measurement sets to the calibrated and flagged state they were at the end of the standard calibration process

- Calibrations are stored as flag columns and calibration tables
  - Might be multiple versions (versions of pipeline or CASA)
- Restore is already in use to recreate calibrated MSs for ALMA and VLASS
- Automated process to deliver to users.
  - Features such as selection and averaging can be applied as part of download process.
- Or as starting point to do optimized imaging etc.
- Special case of only restoring flags will be supported

Use Case: Recalibration

the process of redoing a calibration

- Redo the calibration using
  - Different version of the supporting software tools
  - Additional inputs from the user.
  - Alternative Strategy
- Quality assured recalibration products are stored in the archive
  - Calibrations with non-standard flagging, calibration strategy, or parameters will not be stored in the archive.
- Product available for download, or as input to subsequent processing.

Special Case: Observatory bulk recalibration, fixing an error in the pipeline or updating everything to a better version of the pipeline.

Use Case: Combined Imaging

takes data taken from multiple configurations (including the ALMA Total Power Array) to produce quality assured images with an increased range of emission scales

- Multi-configuration imaging for ALMA and VLA
  - Includes ALMA Total Power
  - Does *NOT* include observations from other telescopes
- Each configuration is calibrated and imaged independently
- When observations complete a combined image is produced
  - Requires changes to project model and tools to support this
- Easiest for observations designed to be used this way
  - Matched SPWs, desired resolution, etc.
  - Do this first
- Eventually want do this with archival data not taken as part of a single project (within reason).

Use Case: Time Critical Observations

modifies the standard processes to support triggered and target of opportunity observations

- For some projects, quick access to products is more important than quality.
- Identify these projects at proposal time
  - There is an extra cost in terms of compute and staff. Should only be used when justified.
- Identified three types of products.
  - Initial Calibration
  - Quick Look Image (shallow automated clean)
  - Standard Image
- May need to be able to preempt the processing queue to get these processed ASAP.
  - Case for minimizing latency, rather than throughput

No Observatory QA before products go to PI!

Use Case: Large & Commensal Projects

maximizing the return to the project team, the Observatory and archive users

- Known groups with experience in radio interferometry.
  - Develop their own QA process and clearly document it.
  - Community products, but limited in scope

- First Phase: Assist projects in making data available
  - Serve [reasonable] products from the NRAO archive
    - Catalogs, Images
  - Provide dedicated search facilities and additional meta-data as requested by the project teams.

- Second Phase: Assist in product creation
  - Customized pipelines, QA assistance, etc…

Later!

Use Case: Curation and Reproducibility

assisting the NRAO user community in documenting reproducible data reduction processes

- Reproducibility of major concern throughout science

- Standard pipelines provide a very controlled and reproducible data reduction path.
  - Inputs (and version) all that is required to reproduce product
  - Allow users to cite pipeline products in journals, addressing open-data standards.

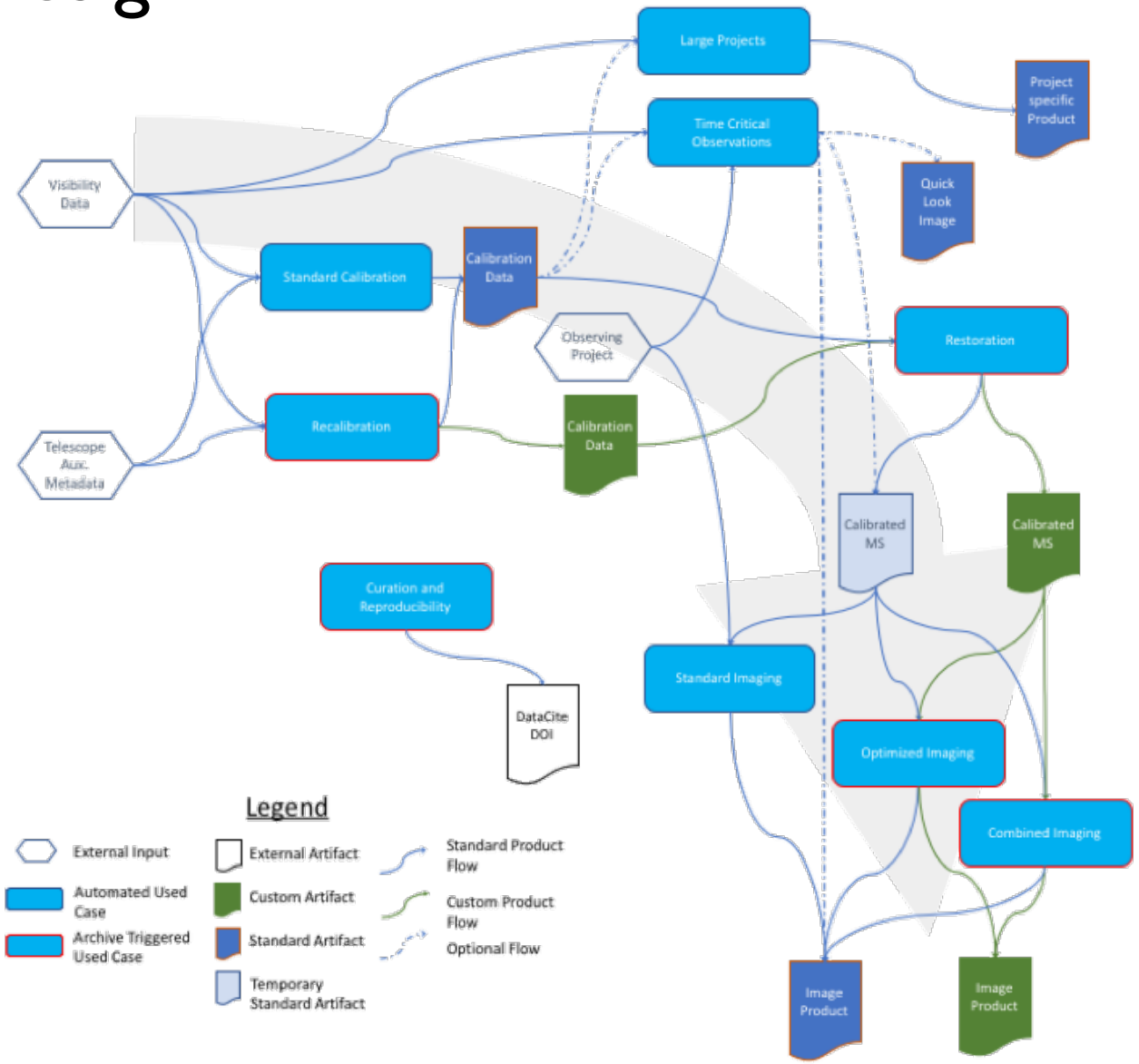
- Minting of DataCite Digital Object Identifiers (DOIs) that will point to a landing page.
  - Similar to capability provided by Mikulski Archive for Space Telescopes (MAST)

Use Case: Commissioning and Validation

special considerations that are necessary for the SRDP staff to commission and validate deliverables

- Set of requirements to allow staff to test and validate products *before* they are put into production.
  - execute SRDP workflows with candidate versions of the software.
  - execute portions of the SRDP workflows
- Due to long running jobs, it will not be possible to bring the system to a fully quiescent state prior to version changes, or updates. It must be possible to modify the system without losing the current execution state, or in such a way that the state information can be recaptured.
- The execution environment may need to be modified, for example using a non-standard destination directory to accumulate outputs from a regression testing run.

System Design



Capability Roadmap

Early-Project

- VLA Science quality Calibration
  - Pipeline Improvements
  - Process changes
- Large Project Ingest
- Input to PST/OPT rework
- QA Definition and Improvement
- Infrastructure and Archive Interface
  - Restore Capability
  - Optimized Imaging (ALMA)

VLASS and ALMA continue to advance pipeline capabilities.

Capability Roadmap

Mid-Project

- VLA Calibration Pipeline Maturing
  - More modes and bands
  - Redesigned PST coming on line with process changes
    - Beginning of Time Critical Use case
- Recalibration workflows
- Begin optimized Imaging for VLA
  - Significant intervention expected
- Minting of DOIs for Users

Capability Roadmap

Late-Project

- Standardizing operations
- Expanded support for Large Projects
- Standard imaging for VLA
- Image Combination

Next Steps

- System Concept
  - Feedback from Staff
    - Presentation in Socorro (6 March)
    - Presentation in Charlottesville (12 March)
  - User Requirements Review (Complete End of March)
    - Users Committee
    - CASA Users Committee
- SRDP Conceptual Design Review 10-11 May

- Starting to develop requirements / schedule for first wave of implementation.

science.nrao.edu
public.nrao.edu
ngvla.nrao.edu

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.