



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A



Computing and Software System Design Description: SDA

020.50.30.00.00-0001 DSN

Status: **RELEASED**

PREPARED BY	ORGANIZATION	SIGNATURE
M. Whitehead, DMS Software Architect	DMS, NRAO	Signed by: Mark Whitehead FAC39E16435B4EE...

APPROVALS	ORGANIZATION	SIGNATURES
R. Rosen, CSS IPT Lead	ngVLA, NRAO	Signed by: <i>Rachel Rosen</i> 4602AA3F0A7C428...
E. Murphy, Project Scientist	ngVLA, NRAO	Signed by: <i>Eric Murphy</i> 82D35C8655C14D1...
J. Kern, AD, DMS	DMS, NRAO	Signed by: <i>J. Kern</i> 014419F0FBC647B...
R. Selina, Project Engineer	ngVLA, NRAO	Signed by: <i>R. Selina</i> 9740299D48964DD...
P. Kotzé, Systems Engineer	ngVLA, NRAO	Signed by: <i>P.P.A. Kotzé</i> 9E9E22E568FB448...

RELEASED BY	ORGANIZATION	SIGNATURE
W. Hojnowski, Project Manager	ngVLA, NRAO	Signed by: <i>William Hojnowski</i> 722DEBE68DE64DB...



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

Change Record

Version	Date	Author	Affected Section(s)	Reason
I	2025-08-11	M. Whitehead	All	Initial draft
A	2025-08-13	M. Archuleta	All	Minor edits and formatting for pdf and release.



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

Table of Contents

I Introduction..... 5

1.1 Purpose..... 5

1.2 Scope 5

1.3 Software Architecture Model..... 6

2 Reference Documents..... 6

3 System Overview..... 6

4 Constraints 8

4.1 Brownfield System..... 8

4.2 Service-based Architecture 8

5 Architecturally Significant Requirements 9

5.1 Availability..... 9

6 Architecture Decision Summary..... 10

7 SDA Conceptual Design 11

7.1 Service-based Architecture 12

7.1.1 SDA-UI..... 12

7.1.2 Archive Services..... 12

7.1.3 Workflow 13

7.1.4 System Integration..... 13

7.2 Technical Infrastructure 13

8 Future Work 14

8.1 Data Architecture 14

8.2 Technical Infrastructure 14

8.3 Deployment Model..... 15

8.4 Test Plan..... 15

8.5 Agile Development Plan 15

9 Appendix A Needs and Requirements..... 16

9.1 Ingest..... 16

9.2 Access..... 19

9.3 Query..... 20

9.4 Retrieve..... 20

9.5 Curate..... 21

9.6 Technical Requirements 21

9.7 Other 22

10 Appendix B Example State Machine 24

11 Acronyms..... 25

List of Figures

Figure 1. General overview of the ngVLA Computing and Software System (CSS). 5

Figure 2. NRAO Science Archive System Container Diagram..... 7



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

Figure 3. SDA Context Diagram..... 11

Figure 4. SDA Container Diagram. 12

Figure 5. Example Capability Service State Machine. 24

List of Tables

Table 1. Availability. 10

Table 2. Architecture Decision Summary. 10

Table 3. Ingest Needs. 16

Table 4. Access Needs. 19

Table 5. Query Needs..... 20

Table 6. Retrieve Needs. 20

Table 7. Curate Needs..... 21

Table 8. Technical Requirements..... 21

Table 9. Other Needs. 22



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

I Introduction

1.1 Purpose

This document describes the ngVLA Science Data Archive (SDA) design constraints, architecturally significant requirement, conceptual design, and plan for evolving the design to the Preliminary Design Review (PDR). SDA is part of the ngVLA Computing and Software System (CSS) shown in Figure 1.

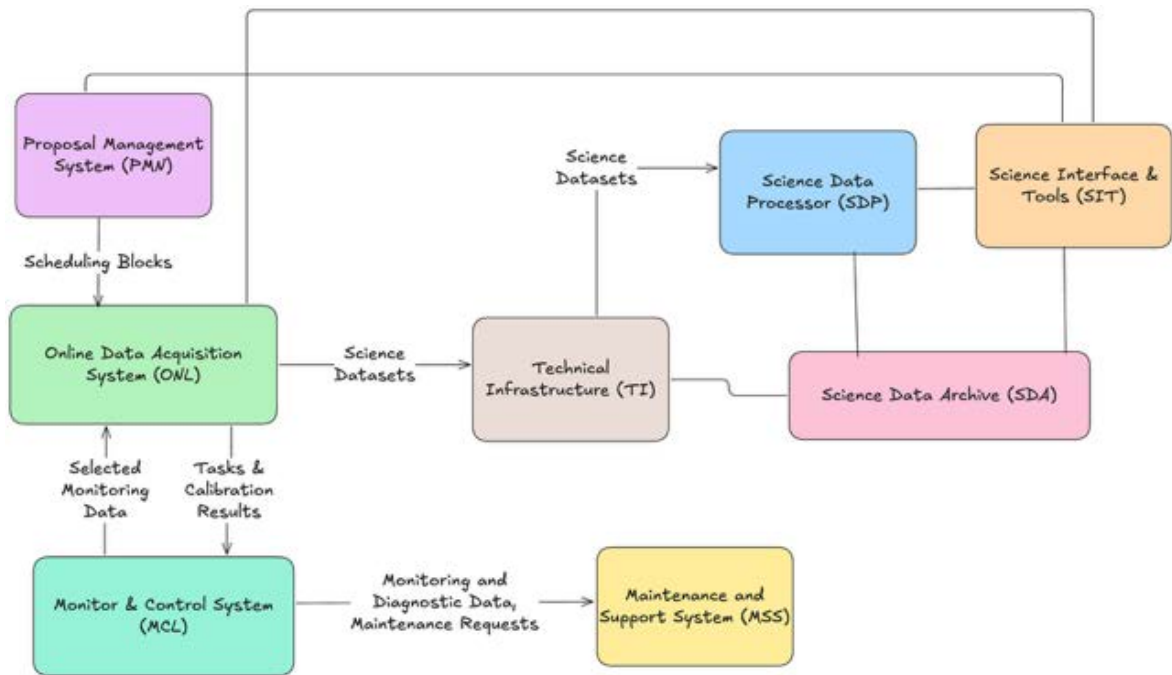


Figure 1. General overview of the ngVLA Computing and Software System (CSS).

1.2 Scope

This document references the current NRAO Science Archive system and highlights the software engineering concerns needed to evolve the current system into the future SDA system on the time scale of ngVLA construction. The backlog of short and intermediate term changes to the current system are not included.

The Technical Infrastructure (TI) system provides and manages computing, storage, and networking resources. While TI details are out of scope for the SDA conceptual design, this document describes the concepts needed to establish separation of concerns and interfaces between TI and SDA.



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

1.3 Software Architecture Model

In this document, C4 is used to model the software architecture¹. C4 is a lean, developer-friendly approach to software architecture diagramming that reflects how software architects and developers think about software. C4 elements include:

- Context: The system that is modeled.
- Container: In C4, a container is not a Docker container. Instead, a container is any process or data store that needs to be running for the overall system to work.
- Component: A grouping of functionality encapsulated behind a well-defined interface.
- Code: Basic building blocks of the programming language; classes, interfaces, functions, etc.

2 Reference Documents

Ref. No.	Document Title	Rev/Doc. No.
RD01	CSS Stakeholder Needs – SDA, SDP, SIT, TI	020.50.00.00.01-0009 REQ
RD02	Radio Astronomy Data Processing System User Interfaces (RADPS-UIs), Concept of Operations, v0.3.	
RD03	ngVLA Software Development Plan	020.50.05.00.00-0009 PLA
RD04	“Summary of ngVLA Processing & Archiving Workflow Concept”, Hibbard, Feb., 2025	

3 System Overview

Figure 2 shows a C4 container model for the current NRAO Science Archive system which encompasses two core functions: archive services and workflow orchestration.

¹ <http://c4model.com>



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

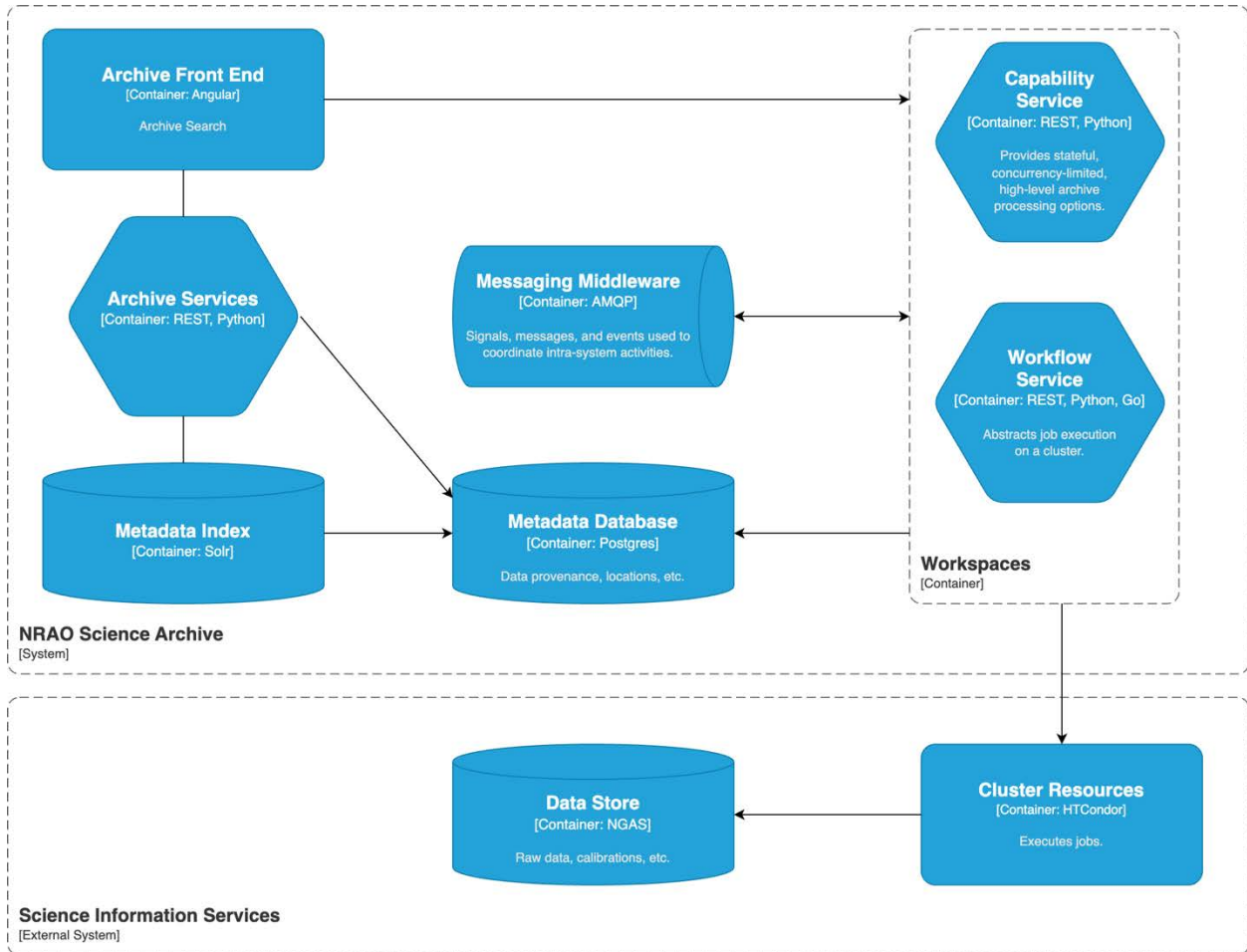


Figure 2. NRAO Science Archive System Container Diagram.

The Archive Front End provides access to Archive Services and to Workspaces, a custom workflow orchestration system that allows users to initiate, version, and track data processing jobs in multiple clusters via HTCondor². The Workspaces system consists of the Capability Service and the Workflow Service.

Generally, archive services support ingestion, search, and retrieval of raw data and data products for the following observatories:

- EVLA and Legacy VLA (Raw data, Calibrations, Standard Images)
- ALMA (Raw data, Calibrations, AUDI Images)
- GBO (Raw data)
- VLBA (Correlated data)
- GMVA (Correlated data)
- Arecibo (Images)

² <https://htcondor.org/>



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

The Capability Service allows NRAO staff to express “business logic”, in this case calibration and imaging workflows, in terms of state machines. An example state machine that supports human interaction is shown in Appendix B. Additionally, the service supports versioning workflows and managing a queue of simultaneous workflows.

The Capability Service delegates workflow execution to the Workflow Service. The Workflow Service manages job submission via an HTCondor instance to DSOC, NAASC, and NMT compute resources. The Workflow Service supports workflow state persistence and recovery which enables scalable, fault-tolerant processing.

4 Constraints

This section describes existing constraints that lead to two key architecture decisions.

4.1 Brownfield System

The current NRAO Science Archive system will be modified to support the ngVLA observatory. The key concerns leading to this decision include cost, operations continuity, and staff experience. Leveraging existing infrastructure and components typically results in significantly lower upfront costs compared to building a new system from scratch. Additionally, modifying the existing archive allows the NRAO to maintain continuous operations with minimal disruption, avoiding the risks and complexities of migrating users, data, and workflows to an entirely new platform. Finally, NRAO staff have a deep understanding of the current system and are already continually and incrementally improving the system design and implementation.

Architecture Decision: The current NRAO Science Archive will be viewed as a brownfield system, a legacy system that is already in place and operational. As such, SDA development will involve modifying, integrating, or extending the current system to incorporate SDA requirements. The scope of the evolution from the current state to the future state includes:

- **Legacy Components:** The system includes frameworks, libraries, or hardware dependencies that may need to be preserved or carefully migrated.
- **Technical Debt:** Historical design decisions and accumulated code complexity may require mitigation to achieve SDA goals.
- **Integration Requirements:** New components or services may need to coexist with and interact with existing ones leading to temporary hybrid architectures.
- **Operational Continuity:** The current archive is a mission-critical system, so changes must be carefully managed to avoid downtime or service disruption.

4.2 Service-based Architecture

Appendix A lists the stakeholder needs which are largely derived from current ALMA and NRAO archive operations [RD01], [RD04]. Given that DMS is currently refactoring the legacy monolithic archive system into services and this approach is an accepted software engineering best practice, the needs are organized by service type. Definitions for each general service type in accordance with the needs are listed below.



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

- **Ingest:** The process of accepting and validating incoming observational data and associated metadata from an observatory's data pipelines and storing them in the archive in a structured, searchable, and standardized format.
- **Access:** A mechanism that controls user access to data based on defined rules or policies such as proprietary periods, user roles (e.g., PI, collaborator, public), or data sensitivity and ensuring that only authorized users can view or download specific datasets.
- **Query:** A structured request submitted by a user or application to search the archive's metadata and catalogs, often using criteria such as sky coordinates, frequency band, observation date, or instrument configuration to locate relevant datasets.
- **Retrieve:** The act of accessing and moving one or more datasets from the archive to a specific location following a successful query, often with options to select data subsets or apply transformations (e.g., averaging, format conversion) to reduce volume or tailor to user needs.
- **Curate:** The long-term management of archived data to ensure its integrity, usability, and scientific value. This includes organizing metadata, reprocessing with improved calibrations, maintaining provenance, and supporting interoperability and future reuse.

Architecture Decision: SDA will utilize a service-based architecture comprised of coarse-grained services that share data stores and infrastructure and are deployed together as a group to support archival use cases. The SDA services decomposition will be organized by requirements related to ingestion, access policies, search queries, retrieval, and curation. Service orchestration or choreography³ will be used as appropriate to accomplish use cases that span multiple services. If archive requirements evolve in a way that requires fine-grained services with highly focused responsibilities that need to be independently deployable, SDA will migrate to a microservices architecture.

5 Architecturally Significant Requirements

An Architecturally Significant Requirement (ASR) is a requirement that has a large impact on the architecture such that changing the requirement would significantly change the architecture. A quality attribute (QA) is a measurable or testable property of a system used to indicate how well it satisfies stakeholder needs.

The general approach for creating the SDA conceptual architecture includes identifying ASRs, organizing ASRs by quality attributes, deriving architecture decisions from ASRs, and using those decisions along with constraints to decompose SDP into subsystems and other entities.

The subsection header below provides the quality attribute, the subsection table summarizes the related ASR, and the subsection text summarizes the key architecture decision.

5.1 Availability

³ Service orchestration involves a central controller explicitly managing interactions among multiple services to complete a use case. Service choreography involves each service interacting with others based on a shared set of rules. Instead of a central controller, services react to events and messages in a loosely coupled manner.



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

Table 1. Availability.

ID	Name	Text
CSS9015	Science Data Archive Product Availability	The DPSS and users need all data products to be available through high availability, high user capacity, low latency archive.

Based on the system requirements, the primary architecturally significant requirement is availability, which for SDA refers to consistent and responsive access to vast volumes of scientific data by a global community of users. Availability is a general system attribute that encompasses several key characteristics related to high availability, high user capacity, and low latency:

- **Fault Tolerance and Redundancy:** The system must withstand hardware failures, network disruptions, or software faults without significant downtime.
- **Scalable Architecture:** To ensure availability under high user load, the system must scale horizontally to handle concurrent queries and data access requests without performance degradation.
- **Fast Recovery and Failover:** In the event of a failure, the system should recover rapidly, ideally with automatic detection and switchover to backup systems, to minimize service interruption.
- **Monitoring and Alerting:** Continuous monitoring of system health, performance, and usage patterns enables proactive detection of anomalies and rapid response to emerging issues, thus maintaining high uptime.
- **No Downtime Maintenance:** Support for rolling updates, hot patches, and live migration helps avoid service interruptions during system maintenance or upgrades.

The next phase of the project will specify availability metrics tailored to the archive's mission in enabling timely scientific analysis and public data access. Other data-related attributes, for example attributes mitigating data loss or governing data retention, will be addressed in the TI design.

Architecture Decision: The primary architecturally significant requirement is availability, which encompasses characteristics related to high availability, high user capacity, and low latency.

6 Architecture Decision Summary

Table 2. Architecture Decision Summary.

ID	Quality Attribute	Description
AD01	N/A (Constraint)	The current NRAO Science Archive will be viewed as a brownfield system and SDA will evolve from it.
AD02	N/A (Constraint)	SDA will utilize a service-based architecture.
AD03	Availability	The primary architecturally significant requirement is availability. Testable statements related to high availability, high user capacity, and low latency will be generated in the next project phase.



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

The current NRAO Science Archive system structure is largely monolithic, uses a shared database model, and reflects the organization of DMS into centralized development teams. The main software engineering goal for SDA is to migrate the monolith to a service-based architecture to achieve a modular design where core business logic is split into discrete services used across applications.

A service-based architecture is suitable for addressing the availability characteristics listed in Section 5.1. However, if SDA eventually requires highly scalable, highly fault-tolerant, independently deployable services then SDA could readily migrate from a service-based architecture to a microservices architecture which would also address availability characteristics. This effort would involve refactoring the shared database model into a decentralized model where each microservice owns its own data.

7 SDA Conceptual Design

Figure 3 shows the ngVLA system context from the perspective of the SDA. Connecting lines represent generic relationships.

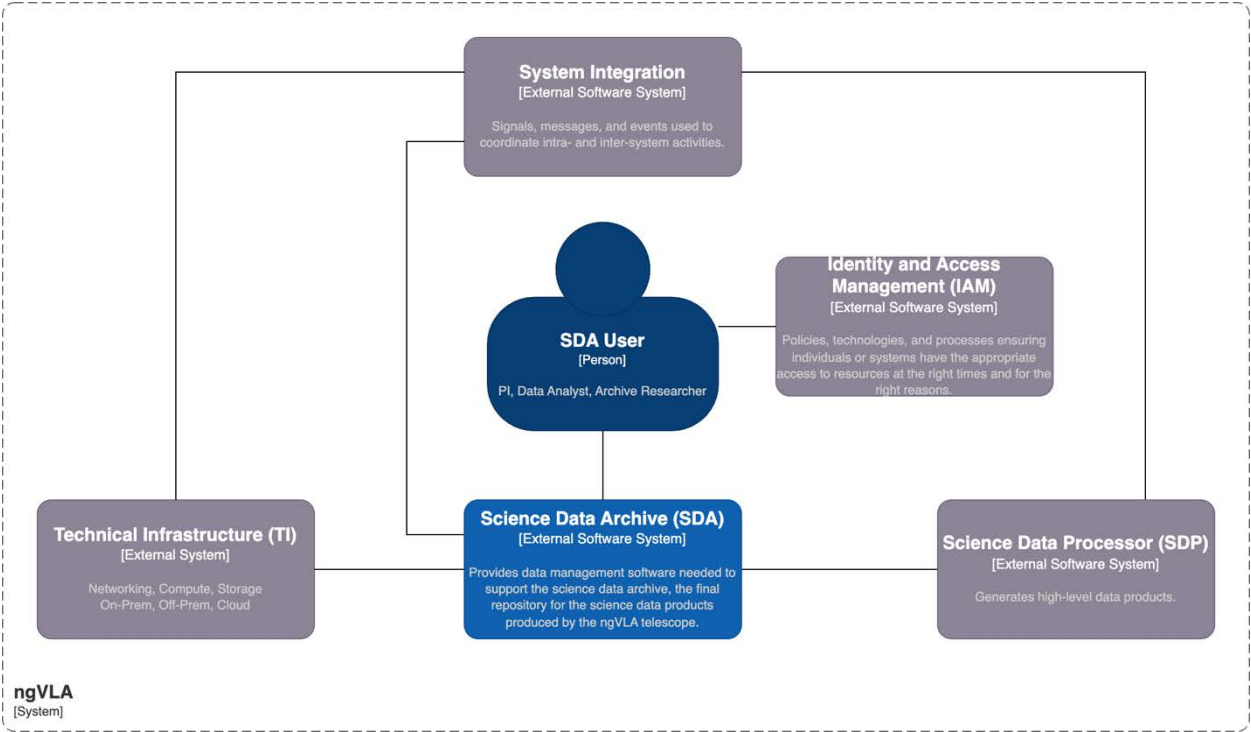


Figure 3. SDA Context Diagram.

From [RD04], the overall vision for SDA is to enable users, including principal investigators and their designees, to be notified when their data becomes available, with direct access links and usage instructions for viewing processing reports, scripts, and data previews using visualization tools that support in-place inspection. During a proprietary period, access will be restricted. Search query features include selecting subsets or averaged data to minimize downloads, retrieving restored calibrated visibilities, and using both interactive and scriptable query interfaces. SDA will support VO standards for interoperability with



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

external survey data, allow saving and sharing of queries, and offer notifications when proprietary data becomes public, including integration with calendar applications.

7.1 Service-based Architecture

Figure 4 shows a C4 container diagram for a notional SDA future state based on AD02.

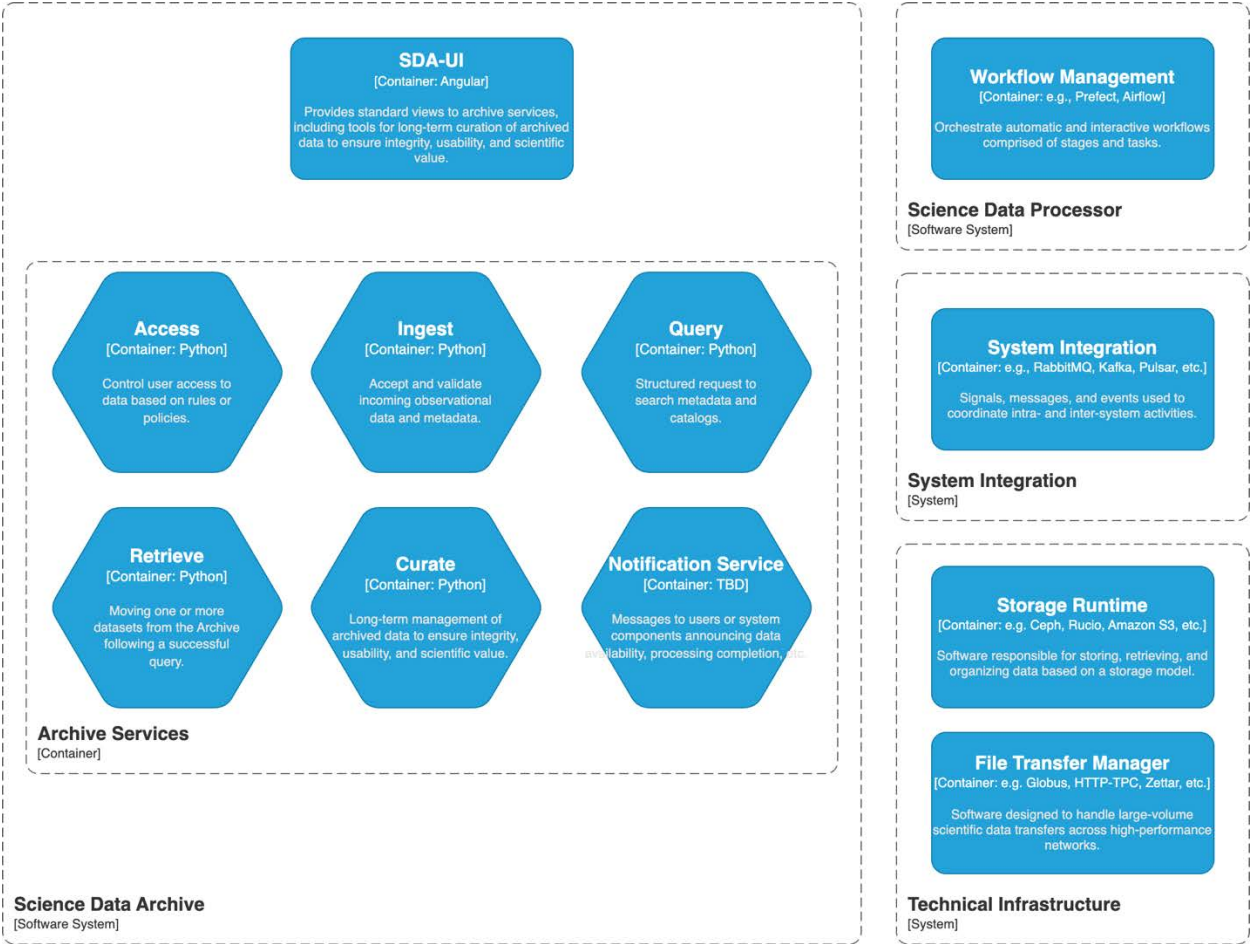


Figure 4. SDA Container Diagram.

7.1.1 SDA-UI

SDA-UI will evolve from the current Archive Front End in accordance with [RD02].

7.1.2 Archive Services

Current incremental development will continue to decompose the current system into services as described in 4.2, including service features that satisfy SDA needs and requirements. This is consistent with the agile approach outlined in [RD03].



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

7.1.3 Workflow

SDA could delegate processing needs to the SDP Workflow Management subsystem. This option would eliminate bespoke Workspaces code in favor of existing commercial-off-the-shelf or open-source products and reuse an ngVLA product that provides the same functionality. Capability state machines could be refactored or translated as appropriate.

7.1.4 System Integration

The current NRAO Science Archive messaging middleware will likely be redundant with ngVLA System Integration components and so SDA could evolve toward using the latter.

7.2 Technical Infrastructure

The Technical Infrastructure system provides and manages computing, storage, and networking resources. In the context of the SDA, Technical Infrastructure also encompasses the concept of a Science DMZ⁴, a specialized network architecture designed to optimize the movement of large scientific data sets between an archive and data processing facilities. The Science DMZ should support:

- **High-Speed Ingestion:** Efficient transfer of raw and calibrated observational data from telescopes or processing centers into the SDA.
- **Data Distribution:** Rapid downloads for users retrieving large datasets or intermediate products for reprocessing or analysis.
- **Workflow Integration:** Data processing workflows that move large volumes of data between the archive and data processing facilities.
- **Isolation:** Separation of operational, user, and archive data traffic, allowing the archive system to function reliably without general-purpose network interference.

SDA-related Technical Infrastructure will incorporate several concepts to manage the storage, organization, and movement of large-scale scientific datasets. The concepts include data storage models, storage systems, storage runtimes, and file transfer management:

- The data storage model will define how data is logically structured. Common models include file, block, and object storage.
- Storage systems provide the underlying physical infrastructure, such as high-speed disks (HSD), high-speed parallel file systems (HSPF), network-attached storage (NAS), storage area networks (SAN), cloud-based storage, and tape systems.
- Physical storage resources are orchestrated by a storage runtime, which is software responsible for storing, retrieving, and organizing data based on the chosen storage model. Examples of such software include Ceph (supporting object, block, and file storage), Rucio (optimized for object and file transfer in scientific environments), and Amazon S3 (object storage).
- To ensure efficient, reliable, and secure data movement between storage systems or institutions, SDA could utilize file transfer managers such as Globus, HTTP-TPC, or Zettar, which are designed to handle large-volume scientific data transfers across high-performance networks.

⁴ <https://fasterdata.es.net/science-dmz/>



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

Since these concepts and potential technical solutions are required for ngVLA, the current archive system will need to be refactored in consonance with the SDA needs and the high availability architecturally significant requirement. The refactoring process provides opportunities to define and enforce clear separation of concerns between Archive Services and Technical Infrastructure via well-defined interfaces.

8 Future Work

This section describes activities to be done between CDR and PDR and choices that need to be made prior to PDR.

8.1 Data Architecture

Data architecture is the high-level design or blueprint for how data is collected, stored, integrated, processed, governed, and accessed. It defines the structure and standards for managing data assets and aligning data-related technology with organizational goals.

The current DMS data architecture consists of traditional on-premises, siloed databases. ngVLA systems will significantly increase the volume, velocity, and variety of data that DMS must manage. Under the known future conditions, the current data architecture will be difficult to scale and maintain and will severely limit opportunities to integrate data to gain operational or scientific insights, implement data-driven features, or increase machine learning-based or other types of automation.

Key components of a data architecture blueprint include:

- **Data Domains and Sources:** Logical grouping of data based on domain (e.g., instruments, commissioning, simulation).
- **Data Models and Schemas:** Definitions of how data is structured and related (e.g., ontologies, entity relationship models).
- **Metadata:** Schemas and keywords defining human- and machine-readable information that describes observational datasets and data products (e.g., provenance, spatial/temporal coverage, instrument configuration).
- **Storage Systems:** Where data lives (e.g., object stores, databases, etc.).
- **Integration:** Flows that connect and transform data across systems.
- **Governance:** Standards for data quality, security, provenance, access, and retention.
- **Access and Consumption:** How data is delivered to users (e.g., APIs, query layers, reports, notebooks).

Between CDR and PDR, DMS should create an ngVLA data architecture that addresses the key components listed above and incorporates the concepts described in 7.2.

8.2 Technical Infrastructure

Prior to PDR, the concepts described in 7.2 should be refined in a way that forms a preliminary design for SDA-related Technical Infrastructure. That design will influence the interface design for Archive Services shown in Figure 4 and help advance the SDA design for PDR.



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

8.3 Deployment Model

Prior to PDR, a deployment model should be created that addresses deployment needs for a service-based architecture.

8.4 Test Plan

A quality attribute (QA) is a measurable or testable property of a system that is used to indicate how well the system satisfies the needs of its stakeholders. A Quality Attribute Scenario (QAS) is an unambiguous way to specify a testable quality attribute. Prior to PDR, QAS's for SDA availability should be developed and incorporated into test planning.

8.5 Agile Development Plan

After CDR, DMS should create an SDA roadmap and product backlog encompassing:

- Features derived from the needs shown in Appendix A.
- Prototyping activities related to evaluating technologies and design options.
- Outputs from 8.1 and 8.2 that identify roadmap milestones or backlog items related to evolving the current NRAO Science Archive.



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

9 Appendix A Needs and Requirements

This section decomposes SDA needs into sub-domains and separates the technical requirements from the needs.

9.1 Ingest

Table 3. Ingest Needs.

ID	Name	Text	Traceability
CSS8007	Ingestion of PIDR products	Final PI-produced SRDP products shall be ingested into the ngVLA archive.	CSS9511.5 [AD7]
CSS8016	External Data Products	The data archive shall have provisions for accepting user-produced data products where those products can be quality assured by the Observatory (such as products from Large projects or Legacy projects). In such circumstances the Observatory will approve the user QA process, not the individual products.	SYS0740 [AD7]
CSS8122	Archive of Data from Test Observations	Telescope Support Scientists need data taken during test observations to be archived similarly to science data.	[AD4]
CSS8130	Storage of Data Processing Results	For standardized test observations, Telescope Support Scientists need the data processing results to be stored.	[AD4]
CSS9013	Storage of Raw Calibration Data	The DPSS and users need the raw data from calibration	[AD7]



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

ID	Name	Text	Traceability
		observations to be archived.	
CSS9014	Storage of Calibration Products	The DPSS need the calibration products and all other metrics needed for calibration from observations of calibrators to be archived and retrievable by the data processing system.	[AD7]
CSS9028	Data Acquired Through MOUs	The DPSS who acquire sensitive observations through MOUs with special partners need these observations to be un-discoverable in the archive.	[AD7]
CSS9067	Storage of Processing Logs	The DPSS need the storage of the processing reports (weblogs) of all processing runs, regardless of success.	[AD7]
CSS9067.1	Storage of Products	At the end of processing and successful QA, the DPSS need the storage of products in the appropriate user archives and observatory databases.	[AD7]
CSS9067.2	Ingestion Error	If there are problems with ingesting/storing the data, the DPSS need the data marked with the appropriate flag and left in place until investigated.	[AD7]
CSS9067.3	Ingestion Success	If there are no problems with ingesting/storing the data, the DPSS need the data automatically	[AD7]



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

ID	Name	Text	Traceability
		moved to the next stage in the lifecycle and update the project status appropriately.	
CSS9075.1	Storage of Tunable Parameters	The DPSS need the modified tunable parameters to be stored and accessible for subsequent NSDR runs.	[AD7]
CSS9081	Storage of User-produced Products	Authenticated users need the ability to upload user-produced products, reports, and quality scores and store them in the archive.	[AD7]
CSS9082.3	Check of PI Products for Upload	The DPSS need an automated check of file names and formats for PI products that are submitted for upload.	[AD7]
CSS9110.1	Ingestion of Calibration Products if QACal=Pass	If QACal is Pass, the DPSS need the archive to ingest the calibration products.	[AD7]
CSS9125	Ingestion of Processing Results from External Entities	The DPSS need to accept the results of jobs processed externally and ingest the relevant data & metadata.	[AD7]
CSS9138	Ingestion of Calibration and Final Products	CSV staff need the calibration and final products to be ingested at the same time.	[AD7]
CSS9139	Ingestion of Incomplete Data Product Packages	CSV staff need the archive to ingest final data product packages that are conformant but incomplete.	[AD7]



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

9.2 Access

Table 4. Access Needs.

ID	Name	Text	Traceability
CSS19519	Proprietary period for time-domain observations	The Science Telescope Support and Diagnostic Group needs the system to accommodate different proprietary period policies for time-domain proposals than for regular proposals. The specific policies are TBD.	[AD2]
CSS9016	Proprietary Period Limits	The DPSS need the archive to have restriction protocols.	[AD7]
CSS9031	Setting Proprietary Periods	The DPSS need the proprietary periods to be automatically set but modifiable (individually or in batch) by qualified staff.	[AD7]
CSS9031.1	Proprietary Period	The proprietary period shall be settable on a per-class, per-project and per-scan basis.	SYS0738 [AD7]
CSS9031.2	Proprietary Data Rights	The archive shall permit the enforcement of a proprietary period for both low-level and high-level data products, permitting public access only after the proprietary period lapses.	SYS0733 [AD7]
CSS9031.3	Proprietary Period Trigger	The proprietary period counter shall start once the data products have undergone any	SYS0743 [AD7]



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

ID	Name	Text	Traceability
		automated or manual quality inspections and are made available to the principal investigator.	

9.3 Query

Table 5. Query Needs.

ID	Name	Text	Traceability
CSS9082.4	Archive Visibility of PI Products	PI products that are stored in the archive need to be discoverable as internally processed data to PIs, DPSS, and Archive Researchers but denoted that it was produced and quality assured by external entities.	[AD7]
CSS9097.1	Archive Query on Scientific Metadata	Archive Researchers need to be able to query data products based on scientific metadata, such as line IDs, source detections, and line fluxes.	[AD7]

9.4 Retrieve

Table 6. Retrieve Needs.

ID	Name	Text	Traceability
CSS9032	Access to Individual Observations	The PIs/Co-Is need immediate access to individual observations as they are calibrated rather than waiting for all observations to be collected and the final products generated.	[AD7]
CSS9033	Sharing of Calibration Scans	The DPSS need calibration scans that	[AD7]



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

ID	Name	Text	Traceability
		can be shared across projects so the data processing system can retrieve and apply shared calibrations to the appropriate projects.	

9.5 Curate

Table 7. Curate Needs.

ID	Name	Text	Traceability
CSS9028.1	Expunging Data Acquired through MOUs	The DPSS who acquire sensitive observations through MOUs with special partners need the ability to expunge the observations from the archive after final products are produced and delivered.	[AD7]
CSS9119	Change Status of Archive Products	The DPSS need to change the status of products in archive, including modifying QA Reports, through an automated system regardless of the product life cycle.	[AD7]
CSS9127	Marking of Archival Products for Bulk Reprocessing	The DPSS need to mark products in the Archive that are scheduled for bulk reprocessing.	[AD7]

9.6 Technical Requirements

Table 8. Technical Requirements.

ID	Name	Text	Traceability
CSS9015	Science Data Archive Product Availability	The DPSS and users need all data products to be available through high availability, high	[AD7]



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

ID	Name	Text	Traceability
		user capacity, low latency archive.	

9.7 Other

Table 9. Other Needs.

ID	Name	Text	Traceability
CSS9036	Lifecycle	The DPSS need the entire software system to have a discrete lifecycle so that the data acquired for any proposal can be monitored.	[AD7]
CSS9065	QA Recommendation	At the end of data processing, the DPSS need to have automatic generation of the QA metrics which are then automatically harvested and used to calculate an overall QA recommendation.	[AD7]
CSS9065.2	Science Quality Score Storage	For some individual products, if the Science Data Processor generates an additional Science Quality Score, the Science Data Processor shall store the Science Quality Score in the Science Data Archive.	[AD7]
CSS9068	QAFinal for SMDR	If the DPSS need automated steps based on final QA scores (QAFinal) for SDRM.	[AD7]
CSS9075	Execution of NSDR after Setting Tunable Parameters	After setting the tunable parameters, the DPSS need to be able to submit NSDR for processing which	[AD7]



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

ID	Name	Text	Traceability
		should be executed automatically and the status visible in the dashboard.	



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

10 Appendix B Example State Machine

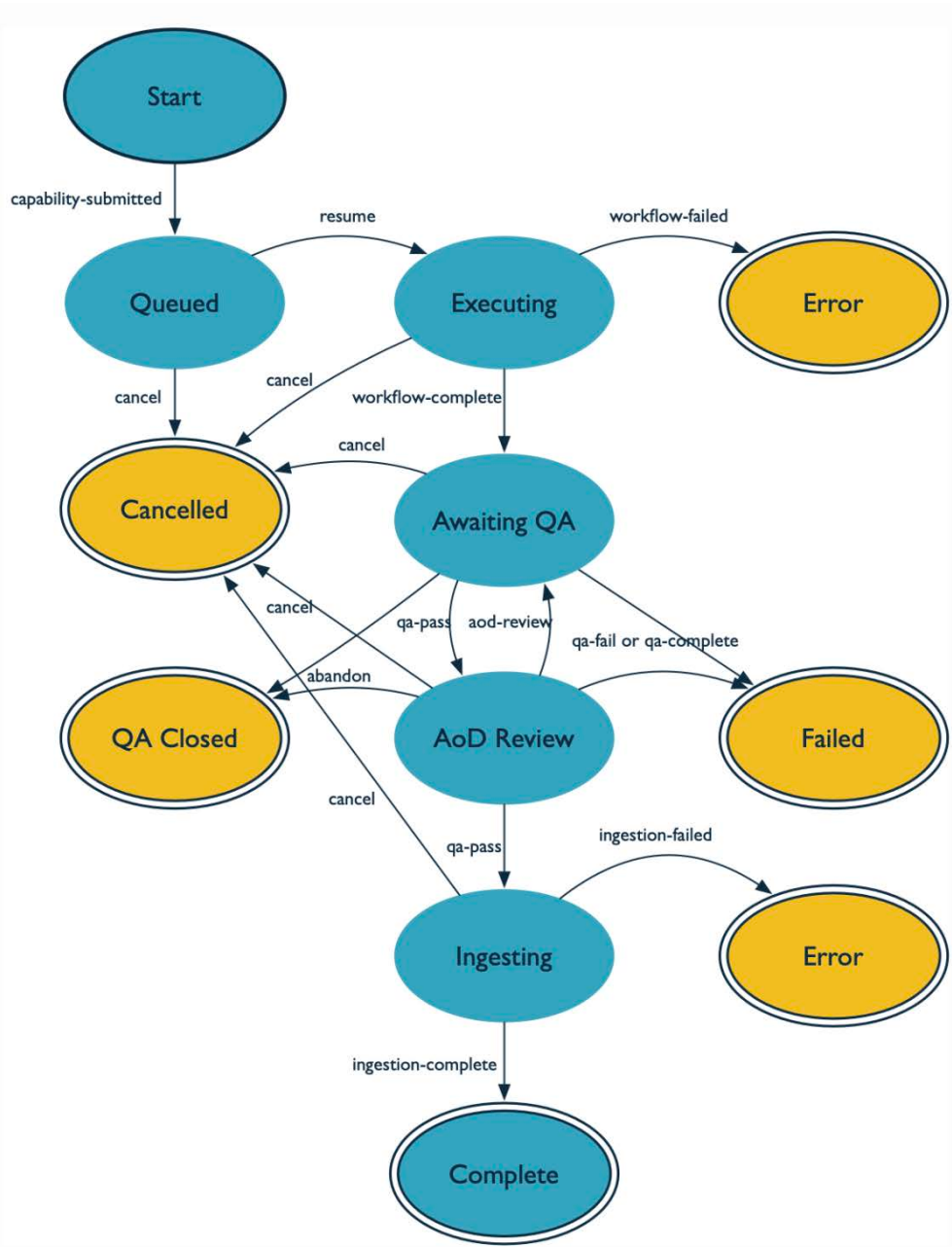


Figure 5. Example Capability Service State Machine.



Title: Computing and Software System Design Description: SDA	Owner: Whitehead	Date: 2025-08-13
NRAO Doc. #: 020.50.30.00.00-0001 DSN		Version: A

II Acronyms

Term	Meaning
AD	Architecture Decision
ALMA	Atacama Large Millimeter/Submillimeter Array
AUI	Associated Universities Incorporated
CSS	Computing and Software System
CSSRR	Computing and Software System Requirements Review
DMS	Data Management and Software Department
DMZ	Demilitarized Zone (networking)
HSD	High-speed disks
HSPF	High-speed parallel file systems
ICD	Interface Control Document
DSOC	Domenici Science Operations Center
IPT	Integrated Product Team
NAASC	North American ALMA Science Center
NAS	Network-attached Storage
ngVLA	The Next Generation Very Large Array Project
NMT	New Mexico Institute of Mining and Technology
NRAO	National Radio Astronomy Observatory
NSF	National Science Foundation
QA	Quality Attribute
QAS	Quality Attribute Scenario
RD	Reference Document
RID	Review Item Discrepancy
SAN	Storage Area Networks
SDA	Science Data Archive
SOW	Statement of Work
TBC	To Be Confirmed
TBD	To Be Determined
VLA	Very Large Array
VLBA	Very Long Baseline Array