

SPACE TELESCOPE RESEACH PROGRAM

Technical Note

TN-2024-01

Observation Data Processing Pipeline v1.0

Authors

David Anderson

Rachel Thompson

November 2024

New York City, New York

Contact:

d.anderson@research.strp.space

r.thompson@research.strp.space

info@mail.strp.space

Executive Summary

Modern astronomical observation programs generate substantial volumes of imaging and metadata that must be processed, validated, and archived before meaningful analysis can occur. As dataset sizes continue to grow, manual processing approaches become increasingly difficult to maintain and scale.

This technical note documents the design and implementation of the first-generation STRP Observation Data Processing Pipeline (ODPP v1.0). The pipeline was developed to support consistent dataset ingestion, automated image processing, metadata organization, and long-term archival preparation.

The primary objective of the project was to establish a reliable and repeatable workflow capable of reducing manual intervention while maintaining data quality standards throughout the processing lifecycle.

Introduction

Observational astronomy relies heavily on the efficient management of large collections of imaging data. Even small observation programs may generate thousands of files requiring organization, verification, enhancement, and archival storage.

Prior to the development of ODPP v1.0, processing activities were largely conducted through independent workflows that varied between projects and personnel. This introduced inconsistencies in dataset organization and increased the effort required for long-term maintenance.

The STRP Observation Data Processing Pipeline was developed to address these challenges through workflow standardization and process automation.

Project Objectives

The project focused on several key objectives:

- Establish a standardized ingestion process for observational datasets.
- Improve consistency in image processing workflows.
- Reduce manual handling of repetitive processing tasks.

- Support metadata preservation and traceability.
- Simplify preparation of datasets for long-term archival storage.
- Provide a foundation for future pipeline development activities.

Pipeline Architecture

The first-generation pipeline consists of four primary processing stages.

Dataset Ingestion

Incoming observational datasets are collected and validated before entering the processing workflow. Basic integrity checks are performed to ensure files are complete and properly organized.

The ingestion stage records essential metadata, including acquisition dates, dataset identifiers, and processing status information.

Image Processing

Following ingestion, image files undergo a series of automated processing procedures designed to improve consistency and facilitate analysis.

Typical processing activities may include:

- Format normalization
- Metadata verification
- Quality assessment
- Preliminary enhancement procedures

The objective of this stage is not to modify scientific results but to prepare datasets for efficient review and analysis.

Metadata Management

Metadata preservation represents a critical component of the pipeline architecture.

The system maintains structured records describing:

- Dataset origin
- Processing history
- Validation status

- Archival readiness

Maintaining comprehensive metadata improves traceability and supports future analytical and operational requirements.

Archival Preparation

The final stage prepares datasets for long-term storage and retrieval.

This process includes organization of supporting documentation, validation records, and metadata packages to ensure future accessibility and reproducibility.

Operational Benefits

Implementation of ODPP v1.0 produced several operational advantages.

Improved Workflow Consistency

Standardized procedures reduced variability between processing activities and simplified project management efforts.

Reduced Manual Processing

Automation of repetitive tasks allowed personnel to focus more heavily on data review and technical analysis.

Enhanced Traceability

The structured recording of processing actions improved transparency and simplified future audits of workflow activity.

Archival Readiness

The integration of metadata management and archival preparation procedures improved long-term dataset organization and accessibility.

Challenges and Lessons Learned

Several implementation challenges were identified during development.

Workflow Flexibility

Observation projects frequently exhibit unique requirements that may not fit perfectly within a standardized process. Maintaining flexibility while preserving consistency remains an ongoing consideration.

Metadata Standardization

Differences in metadata quality between datasets occasionally required additional validation procedures.

Scalability Planning

While ODPP v1.0 successfully supports current program requirements, future growth may require additional automation and infrastructure enhancements.

Future Development

Future pipeline development efforts are expected to focus on:

- Expanded automation capabilities.
- Improved validation and quality assessment procedures.
- Enhanced reporting functionality.
- Integration with future archival systems.
- Support for larger observational datasets.

These improvements will contribute to increased operational efficiency and strengthen long-term data management practices.

Conclusion

The STRP Observation Data Processing Pipeline v1.0 establishes a practical foundation for standardized data management within observational research activities.

By combining dataset ingestion, image processing, metadata management, and archival preparation into a unified workflow, the system improves consistency, transparency, and operational efficiency.

The experience gained during the development of ODPP v1.0 will inform future technical initiatives supporting STRP research and observation programs.

Acknowledgments

The authors thank STRP engineering and data management personnel for their assistance during pipeline testing, workflow evaluation, and documentation review activities.

Correspondence

Software and Pipeline Development:

d.anderson@research.strp.space

Data Management:

r.thompson@research.strp.space

General Information:

info@mail.strp.space

Space Telescope Research Program (STRP)

New York City, New York

United States

© 2024 Space Telescope Research Program