NOAO Mosaic Data Pipeline

Concept Design Review

Frank Valdes
Chris Smith
Rafael Hiriart
Robyn Allsman

NOAO/Data Products Program
Review Committee Members

Tim Abbott            NOAO - South
Andy Becker           LANL
Roc Cutri             IPAC
Daniel Durand         CADC
Ron Probst, Chair     NOAO - North
Buell Jannuzi         NOAO - North
Agenda

Tucson/La Serena

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00 / 12:00</td>
<td>Welcome</td>
<td>D. Shaw</td>
</tr>
<tr>
<td>8:10 / 12:10</td>
<td>Introduction &amp; Charge to the Committee</td>
<td>D. Shaw</td>
</tr>
<tr>
<td>8:30 / 12:30</td>
<td>Project Goals and Requirements</td>
<td>C. Smith</td>
</tr>
<tr>
<td>9:30 / 1:30</td>
<td>Coffee/Lunch Break</td>
<td></td>
</tr>
<tr>
<td>10:00 / 2:00</td>
<td>Technical Presentation</td>
<td>F. Valdes, R. Hiriart, R. Allsman</td>
</tr>
<tr>
<td>11:30 / 3:30</td>
<td>Development Plan</td>
<td>F. Valdes</td>
</tr>
<tr>
<td>11:50 / 3:50</td>
<td>Management Plan</td>
<td>D. Shaw</td>
</tr>
<tr>
<td>12:15 / 4:15</td>
<td>Lunch/Afternoon break</td>
<td></td>
</tr>
<tr>
<td>1:00 / 5:00</td>
<td>Review Panel Executive Session</td>
<td>Review Panel</td>
</tr>
<tr>
<td>3:30 / 7:30</td>
<td>De-Brief for DPP Management</td>
<td>Review Panel</td>
</tr>
</tbody>
</table>
The NOAO Data Products Program is a new program aimed at:

- Identifying scientifically interesting datasets
- Making them available to the community
- Providing tools for data discovery, mining, & exploration

These resources will form a significant component of the NVO. DPP is planning to construct data pipelines for major facility instruments that will generate scientifically important datasets for use by the community.
Programmatic Goals

The programmatic goals for the Mosaic Pipeline project are the following:

- Create scientifically useful data products in the short term, particularly those that enable time-domain science
- Develop in-house expertise in relevant technologies
- Develop experience with routine processing of ground-based imaging data at a high data rate
- Identify observing protocols that will be necessary to support routine processing & DQA for all Mosaic data
- Enable systematic trending of instrument performance
Programmatic Context

The Mosaic pipeline project has explicit or implicit dependencies on the following systems:

- A data handling system (DHS) to capture data from the observing facility. STATUS: operational for several years; upgrades under consideration.
- A data transport system (DTS) to transport data reliably from the observing facilities to the local data center, and then to the mirror site. STATUS: in design, implementation Q1 2003.
- An NOAO Science Archive which provides a data repository, Web-based search capability, & distribution system. STATUS: operational for several months; upgrades under design; formal ingest mechanism TBD.
Programmatic Context – Cont.

Other relevant projects:

- **NEWFIRM instrument pipeline & supporting infrastructure.** The engineering approach will be top-down.
- **LSST Data System Design**
- **Development of a scaleable archive system**
- **Development of an NVO test-bed data portal**
Charge to the Review Panel

Report to the NOAO Data Products Program management your evaluation of the Mosaic Pipeline Project with respect to:

- The utility of the Pipeline for enabling science in the community
- The breadth, depth, and clarity of the requirements specification
- The viability of the concept design
- The feasibility of the implementation plan
NOAO MOSAIC Pipeline
Motivation & Requirements

NOAO MOSAIC Pipeline Team
Introduction

• MOSAIC Imagers are the most heavily used survey instruments at NOAO
• MOSAIC images form the most uniform datasets at NOAO (minimal N/S variation)
• MOSAIC images are arguably the most scientifically useful archive products produced by any currently available NOAO instrument (wide-field imaging)
Project Motivation: Bottom Line

• Produce high quality MOSAIC data products from both Survey and PI-driven observations (for projects and archive)
  – Note that high quality data products require high quality data!

• Support existing and new time-domain science experiments; prepare U.S. community for era of LSST
Project Motivation: Science

• Supporting Existing Surveys
  – Original expectation that data was to be archived
  – NOAO now taking on the responsibility for archiving, in cooperation with survey teams

• Supporting Individual PIs

• Enabling Archive science with uniform Mosaic dataset

• Enabling “survey science” by smaller teams on faster timescales, specifically in time-domain
Science Use Cases: Examples

• Survey Examples:
  – Fundamental Plane Survey
  – The $w$ Project Supernova Search

• Archive and Time-domain Examples:
  – Search for distant supernovae in deep MOSAIC fields
  – Long term variability of AGN in deep extragalactic fields
Project Motivation: Observatory

- Monitor camera status/instrument health
- Monitor telescope status
- Make support scientists more efficient and more effective (important as both KP and CT have fewer FTEs to support the MOSAIC imagers)
- Bottom line: Make sure MOSIAC imagers produce high quality data
Project Motivation: Data Products

• Experience and foundation for future pipeline software efforts
• Exploration of concepts for LSST data processing and data products (both standard and time-domain)
  – e.g., catalogs, classification, visualization
• Experience in pipeline operations for future instruments and projects
Context: Hardware

- NOAO 4m telescopes
  - CTIO Blanco 4m
  - KPNO Mayall 4m
  - (also WIYN 0.9m)
- Mosaic 8Kx8K Imagers (N and S)
  - 4x2 array of 2048x4096 CCDs
- Fast network links to downtown sites
Examples of Instrumental Signatures

Pupil Pattern  Fringing  Crosstalk
Context: Software

• Existing projects/infrastructure:
  – Arcon, DHS, DTS
  – Pipeline *(this project)*
  – Archive

• Additional projects (Pipeline should support):
  – Autologging DB, Environmental DB,
  – Error reporting and tracking DB
  – Integrated Observatory operations DB
NOAO 4m telescopes are classically scheduled.

Traditionally observers define calibration and observation strategies and execution. Political/Sociological change needed to standardize MOSAIC observing practices.

Calibrations? Yes

Canned calibration sequences in afternoon or morning.

Standards? Possibly

e.g., require observations of set fields every night.

Objects? Perhaps

e.g. require use of mosdither or mosocs.
Pipeline Project: Two Phases

- Phase 1: MOSAIC data reduction
  - Ensure data quality
  - Remove instrumental signatures
- Phase 2: Time-Domain Data Products
  - Support of new time-domain projects, e.g., identification of transient objects

- Phase 1 described in detail in this CoDR
- Phase 1 is a prerequisite for Phase 2
Phase 1 Requirements

• Provide DQA in near-real-time
  – Alert on daytime calibration variations
  – Alert on bias variations
    • Both in calibrations and object frames
  – Alert on noise variations
  – Alert on image quality variations
    • Focus, guider problems, etc.

• Time requirements
  – Quick reduction in <~5 min
Phase 1 Requirements (cont.)

• Provide database for investigating long term trends
  – Photometric variations (throughput, color)
  – Astrometric variations
  – Site variations (seeing, etc.)
  – Other instrumental variations
    • e.g., noise characteristics, gain, linearity
  – Combine with other instrument measurements
    • e.g., shutter timing, CCD QE, filter transmission
Phase 1 Requirements (cont.)

- Remove Instrumental Signature from Mosaic data
  - Flat field to <1% (goal <~0.5% in final reductions)
  - Remove fringing to <1%
  - Remove pupil image to <1%

- Astrometric requirements
  - To be limited by USNO-B catalog (~0.2")
  - May use UCAC when/where available

- Photometric requirements
  - Absolute limited by USNO-B catalog (~0.3 mag)
  - Relative to ~5% in quick reduction, ~1% in 2nd pass
Phase 1 Requirements (end)

- Work with instrument scientists, telescope scientists, and NOAO scientific staff to define strategies to meet data quality requirements
  - Required daytime calibration frames
    - Both short term and long term goals
      - Short term: flat and bias sequences, acceptable variations
      - Long term: gain and linearity tests for each observing block
  - Required nighttime strategies
    - Astrometric calibration fields for internal calibrations
    - Photometric calibration, color terms, etc.
    - Observing strategies to encourage/ensure data uniformity
Phase 2 General Description

• Automatically produce and archive higher level data products from standard observations
  – Combined dither stacks
  – Catalogs of objects (stars, galaxies)

• Produce time-domain data products where feasible and/or desired
  – Light curves of stationary objects
  – Detection of moving objects, trajectories

• Define/develop archive structures to support catalogs and support time-domain science
Requirements: Phase 2

- Support automatic combination of dither sequences to produce deep images
- Support transient detection to \( \sim \) poisson limit
- Support automatic generation of catalogs of objects and object properties, including time variations (light curves)
- Support detection and classification of moving objects
Coffee/Lunch Break

• Discussion...

• Coffee!!!

• Discussion...

• Technical Presentation
  @ 10am Tucson / 2pm Serena