National Optical Astronomy Observatory

The NEWFIRM Data Handling System and
Data Processing System
Development Plan

1. Description and Summary

The intent of this note is to describe a plan for development of the NEWFIRM Data Handling System and the NEWFIRM Data Processing System. The specifics for each system are delineated in separate documents, e.g., The NEWFIRM Data Handling System and The NEWFIRM Data Processing System. In summary, the DHS is responsible for capturing the data from MONSOON and the Observatory Control System (OCS), for staging this data for transfer to a data processing facility, for staging this data for on-line processing (Quick Look/Quick Reduce/Astronomer's Sandbox), and for ensuring that the data exchange is complete and correct.

The Data Processing System provides the on-line (real-time and near real-time) processing on the mountain in support of exposure evaluation (Quick Look/Quick Reduce – including all interfaces/GUIs) and in support of the observers analysis (the Astronomer's Sandbox).

The interface between the DHS and DPS is in the form of buffered data that is staged for processing and of files in support of the Astronomer's Sandbox. The following discussion follows this separation.

2. The NEWFIRM DHS Development Plan

The DHS consists of data source interfaces, internal functionality, and data sink interfaces. The data source interfaces defines, implements, and deploys the data capture interface to MONSOON and the OCS. The internal functionality supports correct, complete, and timely transfer between the data sources and the DHS (design, implementation, and deployment inclusive) and between the DHS and the data sinks (DPS and data processing center). The data sink interfaces define, implement, and deploy the interface to the DPS and data processing center.

The NEWFIRM DHS will be implemented in multiple phases as described below.

2.1 Phase 1 - Completion 2/4/04 API, 2/18/04 Implementation

Summary: Transferring PAN pixel and AV data via current NEWFIRM network infrastructure.

The first phase supports analysis of NEWFIRM engineering array data on a machine separate from the PAN. (this is distinguished from the ability to evaluate a PAN generated fits file on the PAN - this capability is being developed by MONSOON).

This is dependent on establishing the interface API. The status of the API is that the work is ongoing and should be finalized by 2/4/04. Implementation will take two weeks following completion of API definition. The internal DHS functionality (as defined in the DHS documentation) will concentrate on a correct transfer (functionality associated with completeness is accomplished in Phase 2).

Throughput is dependent on the network structure between commuicating machines (PAN to Analysis platform). Current installations support 100Mb networks. Additional capability (i.e., additional Network Interface Card(s) and high speed wiring) will be required to be purchased and installed to provide performance that is needed for NEWFIRM.

Deliverable: Code to provide the above capability and associated design/implementation documentation.

2.2 Phase 2 - Completion 4/2/04
Summary: Completion of DHS validation and recovery development. This is a two month effort following Phase 1.

This Phase concentrates on implementing the DHS internal functionality as described above.

Deliverable: Code to provide the above capability and associated design/implementation documentation.

2.3 Phase 3 - Completion 6/4/04
Summary: Full DHS capability as defined in the document currently being developed, i.e., The NEWFIRM Data Handling System.

This involves safe store of data, staging data for archive, forking data to the on-line pipeline, and supporting an astronomer's sandbox on the mountain.

Deliverable: Code to provide the above capability and associated design/implementation documentation.

The above estimates are based on a number of assumptions:

a. That the requirements as currently understood remain stable, e.g., the phase 3 requirement of staging data for transfer to a data processing facility is based on an assumed capability (that there exists a transport facility to manage the functionality associated with providing a safe-store – replicating the data within the transport system to ensure against data loss and providing a release message when the DHS can safely remove its copy).

b. Forking of data to the on-line pipeline (the functionality of which is the subject of section 3) and for the astronomers sandbox is constrained by straightforward development of data transfer by buffer and/or file.

c. Sufficient manpower is available to support development.

d. Hardware is procured in a timely manner to support development.

e. Interface development for Phases 2 and 3 are tied to development of MONSOON and the OCS. Sufficient lead time has been assumed to support this schedule.

A deviation for these assumptions may have an impact on the scheduled dates.

3. The NEWFIRM DPS Development Plan.

The NEWFIRM DPS development is realized by two activities, i.e., developing the fundamental infrastructure for data processing and developing the specific software to accomplish the on-line and science data processing. To the maximum extent possible, the software being developed for the on-line processing pipeline and for the off- line, science processing (assumed to be accomplished at an off-site data processing center) will be shared. Infrastructure refers to all the functionality required to make a pipeline work, e.g., data management, process management, and resource management. Science processing refers to the software modules that are used to create data products.

3.1 NEWFIRM DPS WBS – Completion 5/7/04
The initial effort toward developing the NEWFIRM data processing system (DPS) will be to review existing pipelines to ascertain the extent of development that will be required. This review will concentrate on two aspects of pipeline development, i.e., infrastructure and science software. Elemental

to this study will be the current DPP Mosaic pipeline development effort. It is the intent of this review that all relevant, existing components will be considered (e.g. both IRAF and NASA Goddard s/w provide utilities for file manipulation and we will ensure that duplicated effort is avoided).

Following this review, a plan and WBS will be established, delineating the tasks that need to be accomplished, their relative timing, associated manpower, and a schedule. The following is a listing of tasks that support development of the NEWFIRM DPS WBS:

1. Infrastructure Review

2. Science Software Review

3. Develop NEWFIRM DPS work breakdown, manpower requirements, schedule, deliverables, and milestones.

  a. tasks

  b. manpower

  c. schedule

  d. milestones/deliverables

3.2 DPS Infrastructure – Alpha Release 12/31/04
Summary: This task represents design, development, implementation, documentation, and review of an alpha or initial release of the DPS infrastructure. An Alpha release is characterized as a functionally complete prototype of the DPS (complete yet lacking the robust, fidelity, and performance required).

3.3 DPS Science Software – Alpha Release 12/31/04
Summary: This task represents design, development, implementation, documentation, and review of an alpha or initial release of the DPS science software, with emphasis on the functionality required for on-line processing (QL/QR/AS).

3.4 DPS Infrastructure – Beta Release 4/05/05
Summary: A Beta release represents is increased functionality as compared to the Alpha release. That is there is the functional evolution associated with the incremental design/implementation process as well as incorporation of knowledge gained from testing the Alpha release. It is expected that the Beta release would support mountain based testing.

3.5 DPS Science Software – Beta Release 4/05/05
The Beta release for the DPS Science Software.

3.6 DPS Infrastructure – Production Release 6/30/05
Summary: This task represents refinement, implementation, documentation, and review of the production version of the DPS infrastructure.

3.7 DPS Science Software – Production Release 6/30/05
Summary: This task represents refinement, implementation, documentation, and review of the final version of the DPS science software, with emphasis on the functionality required for on-line processing (QL/QR/AS).

DRAFT - NEWFIRM DHS/DPS Development Plan

Table 1

Task	Completion Date	Deliverable (all releases include documentation + implementation)
DHS		
Phase 1	2/4/04	Phase 1 Release
Phase 2	4/2/04	Phase 2 Release
Phase 3	8/3/04	Phase 3 Release
DPS		
DPS WBS	5/7/04	WBS documentation
DPS Alpha Rel.	12/31/04	Alpha Release - Infrastructure
		Alpha Release – Science S/W
DPS Beta Rel.	4/1/05	Beta Release – Infrastructure
		Beta Release – Science S/W (QL/QR/AS)
DPS Prod. Rel.	6/30/05	Production Release - Infrastructure
		Production Release - Infrastructure