The ESO Data Flow System

Michèle Péron
Software Development Division
ESO
.... what has changed since its inception in 1995 and what has remained the same
Talk Outline

Introduction
Submission, Evaluation and Scheduling of Observing Proposals
Preparation, Scheduling and Execution of Observations
Pipeline (Data Reduction)
Data Archiving and Distribution
The Data Flow System for an ELT

VLT end-to-end operation model
The development of the VLT end-to-end operation model and the requirements analysis of the software started in the fall of 1995 using the Object Modeling Technique (OMT) developed by Rumbaugh around 1991.

A first prototype of the system was verified and validated during the NTT big-bang in 1997. The first release of the system was used during the VLT first light in 1998.

- Few central observation concepts: Observation Block, Reduction Block
- Few Design choices: thin interface to control software, instrument independent applications

Since that time, the DFS has evolved to accommodate:
- Changes and improvements of the operation model
- New instruments (data volume & complexity)
- User requests for better services
- New technologies
The ESO Data Flow System

Formal model describing the system which handles the flow of science data associated with the operation of the Observatory. Focus is on conceptual rather than implementation issues.

Observation Block (OB): smallest observational unit, with a set of correlated exposures and one target.

Object Model (OMT)
The ESO Data Flow System (cont)

Observing Proposals are invited twice a year
The OPC evaluates and grades the proposals
Successful proposals are scheduled
Observation Blocks are prepared, submitted to ESO and validated
Observation Blocks are executed
Resulting data is archived
Resulting data is processed for the purpose of quality control
Program Handling

Introduction
Submission, Evaluation and Scheduling of Observing Proposals
Preparation, Scheduling and Execution of Observations
Pipeline (Data Reduction)
Data Archiving and Distribution
The Data Flow System for an E-ELT
Latex packages are downloaded from an ftp server

Users fill in the Latex form
Users submit the form to ESO per email
The system parses the LATEX form and returns errors in an email
If all is fine, email is sent to request submission of pictures

2011

Users log-in into the User Portal and download the Latex package

Users fill in the Latex form
Users upload the form to ESO through a WEB interface
Pictures can also be uploaded through the WEB.
A PDF file is generated by the system and checked by users
Users submit the proposal.
Program Handling (Long-Term Scheduling)

Scheduling of Observations for an observing period of 6 months. GUI and a constraint programming engine taking in account the constraints of the recommended programs.
Observation Handling

Program Handling

Quality Control

Pipeline (Data Reduction)

Control System

Science Archive

Introduction
Submission, Evaluation and Scheduling of Observing Proposals

Preparation, Scheduling and Execution of Observations

Pipeline (Data Reduction)
Data Archiving and Distribution
The Data Flow System for an E-ELT

ADASS 2011 – Michèle Peron
Observation Handling (OB preparation)
Observation Handling (OB Preparation)

P2PP (2011)
Survey Telescopes (i.e., VISTA, VST) brought in new ways of observing.
One program might span several years and including hundreds of OBs.
Scheduling containers allow astronomers to express more complex strategies by creating additional abstraction on top of individual OBs that allow expressing dependencies between them.
Observation Handling (OB Execution)

Large number of OBs of short duration, with execution dependencies expressed in scheduling containers.

- Ranking engine suggests the next OB to be executed, taking into account weather condition, visibility constraints, user priority, group score as well as the observing run rank.
Science Archive

Program Handling

Quality Control

Pipeline (Data Reduction)

Observation Handling

Control System

Science Archive

Introduction
Submission, Evaluation and Scheduling of Observing Proposals
Preparation, Scheduling and Execution of Observations
Pipeline (Data Reduction)
Data Archiving and Distribution
The Data Flow System for an E-ELT
Data Flow Back End

VLT Instruments → On Line Archive System → Data Transfer System

Quick Look Pipeline

Publish-subscribe model

Primary Archive (NGAS) → Quality Control Process

2011

1998
Since middle of 2008 all VLT/VLTI data are transferred to ESO Garching through the network.

Highly optimized utilization of high-latency network.

File Transfer can be flexibly prioritized.

This new system has enabled “more” Quality Control to take place in Garching.
Data Distribution (request Handler)

Code re-use from CADC/ALMA

Nathalie Fourniol: News about ESO Archive services
Ignacio Vera: hFits: from storing metadata to publishing ESO data

ADASS 2011 – Michèle Peron
Pipeline (Data Reduction)

Program Handling

Observation Handling

Control System

Science Archive

Quality Control

Pipeline (Data Reduction)

Introduction
Submission, Evaluation and Scheduling of Observing Proposals
Preparation, Scheduling and Execution of Observations
Pipeline (Data Reduction)
Data Archiving and Distribution
The Data Flow System for an E-ELT
Data Reduction at the Telescope

Required to control the health of the instruments and check the quality of the observations
Must be done automatically and in quasi real-time
Large amount of data (few hundreds of Gigabytes per night)
Multi-core hardware and parallel processing
Complex instruments and complex reduction algorithms
Pipeline Infrastructure

Raw Data → Data Organiser → Reduction Blocks

OCA Rules

Calibration Files

Reduction Block Scheduler → Pipeline Recipe → Reduced Data
Data Reduction at the Telescope (cont)

Automatic Data Organization (available in 1998 in C++, re-engineered in Java in 2005)

Based on a flexible rule-engine & a domain-specific language

Creates Reduction Blocks (contains all information for reducing a set of related data)

Who am I?
Which data belongs to my group?
Which type of calibration are needed to process me??
Data Reduction at the Telescope (cont)

Reduction Block Scheduler (available in 1998 in C++, re-engineered in Java in 2007)

- Multi-threaded application
- Takes in account dependencies between Reduction Blocks
Pipeline Algorithms – New approaches

Wavelength calibration of a MOS exposure using first guess model to find reference lines

New approaches (such as pattern-matching) are needed

In Memoriam Carlo Izzo
Data Reduction at Home (Reflex)

UVES Workflow For Point Source Echelle Data (v. 4.9.0)

Workflow Instructions

To run this workflow on the demo data:
- Turn on highlighting. Choose "Tools" => "Animate at Runtime" from the menu and set it to "2".
- Press the "Run" button or click it to start the workflow.

To run on a different data set:
- Click on ROOT_DATA_DIR and set as appropriate. All subdirectories of RAW_DATA_DIR will be searched for data.
- If desired, change END_PRODUCTS_DIR.
- Press the "Run" button or click it to start the workflow.

To monitor the progress of the workflow in more detail:
- Open "Window" => "Runtime Window" in the menu before starting the workflow.

Setup Directories

Working Directories:
- BOOKKEEPING_DIR: RSD000_DATA_DIR/reflex_bookkeeping/enu
- LOGS_DIR: RSD000_DATA_DIR/reflex_logs/enu
- TMP_PRODUCTS_DIR: RSD000_DATA_DIR/reflex_tmp/products/enu
- OUTPUT_DIR: RSD000_DATA_DIR/reflex/end_products

Global Parameters

- FITS_HEADER: .fits files to use for the inspection of input/output products
- ESOR discussion -- suppress--verbose=TRUE esorex arguments
- Greedier: N/A
- Change "DataDir" to that to which BOOKKEEPING_DIR, TMP_PRODUCTS_DIR and LOGS_DIR each time the workflow is run (easy mode won't work anymore)
- NRO_DAOPHOT_DIR: 2012-02-14/12/30/001/0000/0000
- GLOBAL_TIMESTAMP: 2013-12-14/13:13:02:20

Step 1: Data Organisation and Selection

Step 2: Creation of Master Calibration Files

Step 3: Wavelength and Response Calibration

Step 4: Spectrum Extraction

Step 5: Output Organisation
Data Reduction at home (Reflex)

Demo: Ballester et al.
Phase 3: Handling Survey Data Products

Phase 3 denotes the process in which principal investigators of ESO observing programmes return their reduced data products to ESO for storage in the ESO archive and subsequent data publication to the scientific community.

The new Phase 3 infrastructure supports the reception, validation and publication of data products from the public survey projects and large programmes to the ESO Science Archive Facility.

J. Retzlaff, M. Arnaboldi, V. Forchi, P. Nunes, S. Zampieri, T. Bierwirth, M. Ron, M. Romaniello, J. Lockhart, D. Suchar (ESO)

http://www.eso.org/sci/observing/phase3.html
**Phase 3 data flow & infrastructure**

The release manager is a web application that allows the P.I. to define data collections and releases and to manage the Phase 3 delegation to co-investigators.

The data is transferred by the PI/Co-I via FTP to the dedicated staging area.

The release validator is a command-line application that helps to verify the data standard and validity of the header keywords against predefined rules.

Interfaces between the Phase 3 data flow, its users and the ESO Science Archive Facility.

Start of operations: 10 March 2011

http://archive.eso.org/wdb/wdb/adp/phase3_main/form

ADASS 2011 – Michèle Peron
End-To-End Operational Model similar to VLT.
Instruments will have the same level of complexity as the VLT ones.
There will not be a DFS-VLT and a DFS-ELT but a DFS
Will be an evolution of the current system or a revolution?

Underlying Data Model has grown organically and might require complete re-engineering
THANK YOU

… to all those who have contributed to the DFS over the past years:

Software Development Division

Data Management and Operations Division

The Observing Programme Office

Science Operations Department of the Observatory