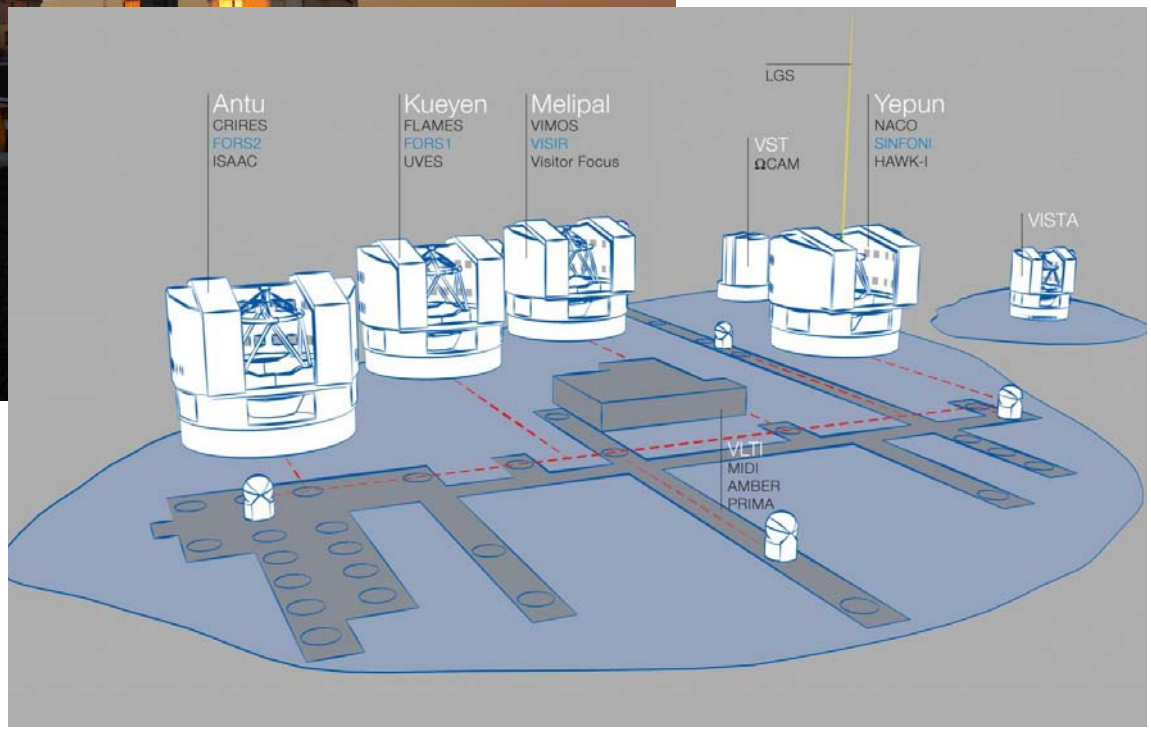
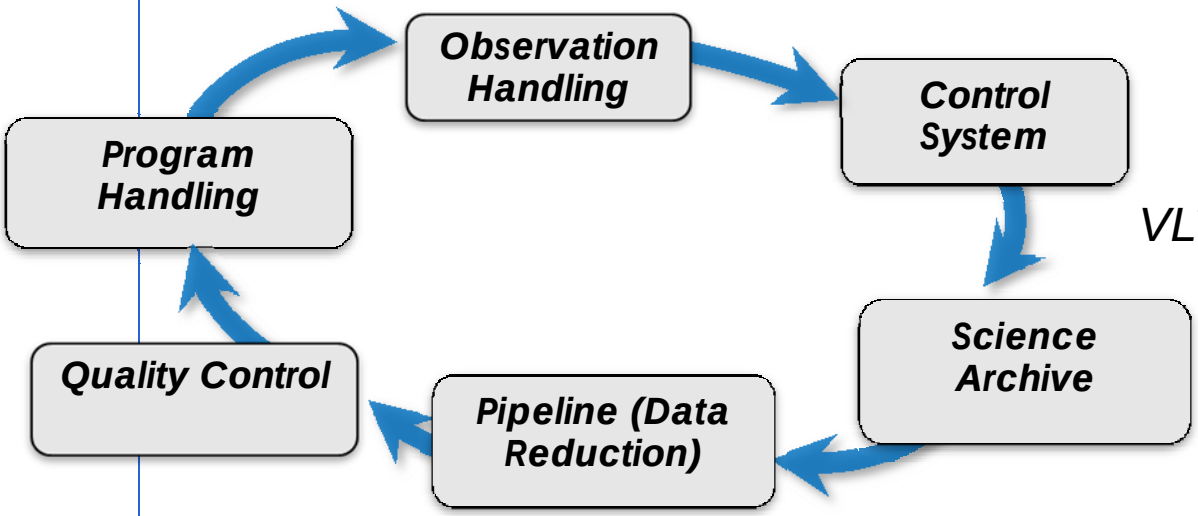
The ESO Data Flow System

Michèle Péron
Software Development Division
ESO

…. what has changed since its inception in 1995 and what has remained the same



Talk Outline



VLT end-to-end operation model

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

The inception of the ESO Data Flow System

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 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:

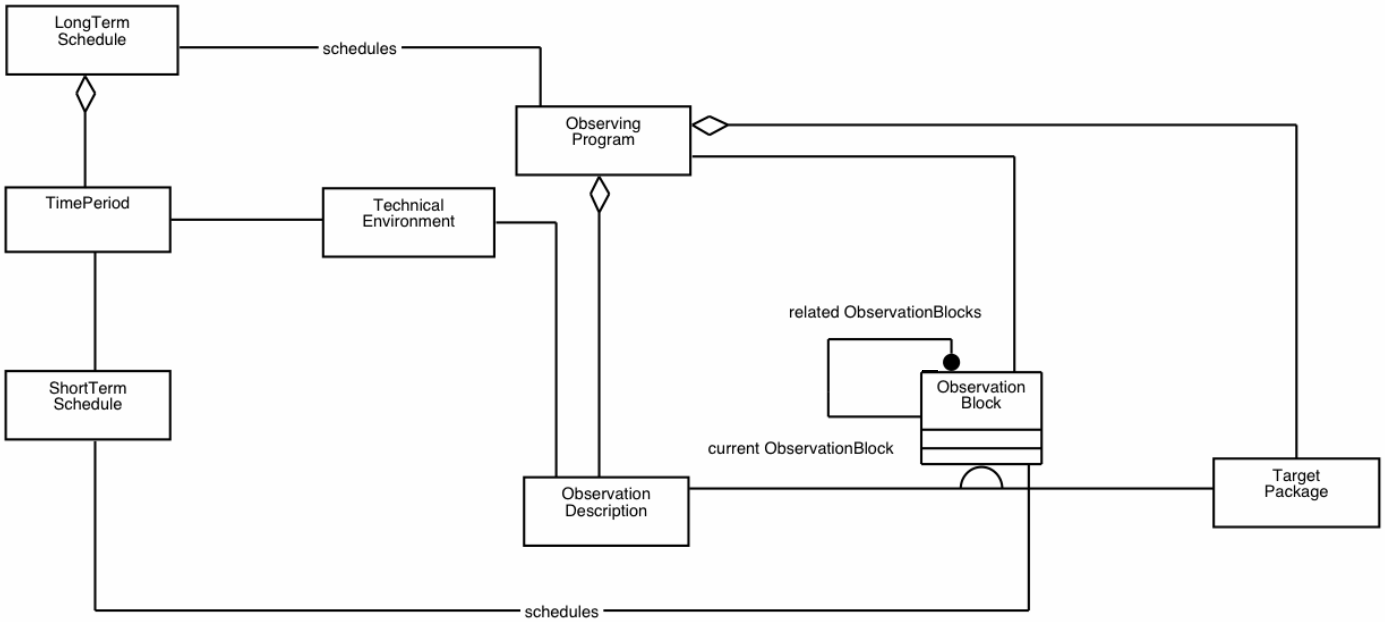
- Changes and improvements of the operation model
- New instruments (data volume & complexity)
- User requests for better services
- New technologies

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

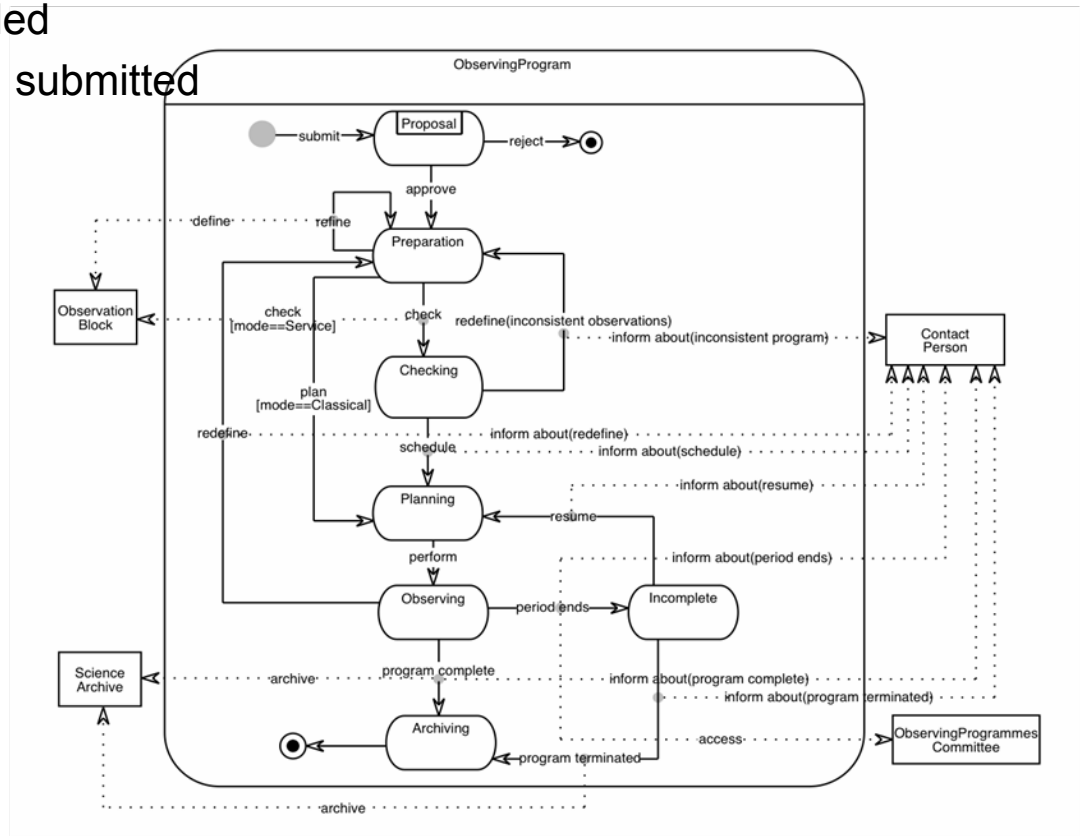
Observation Blocks are prepared, submitted to ESO and validated

Observation Blocks are executed

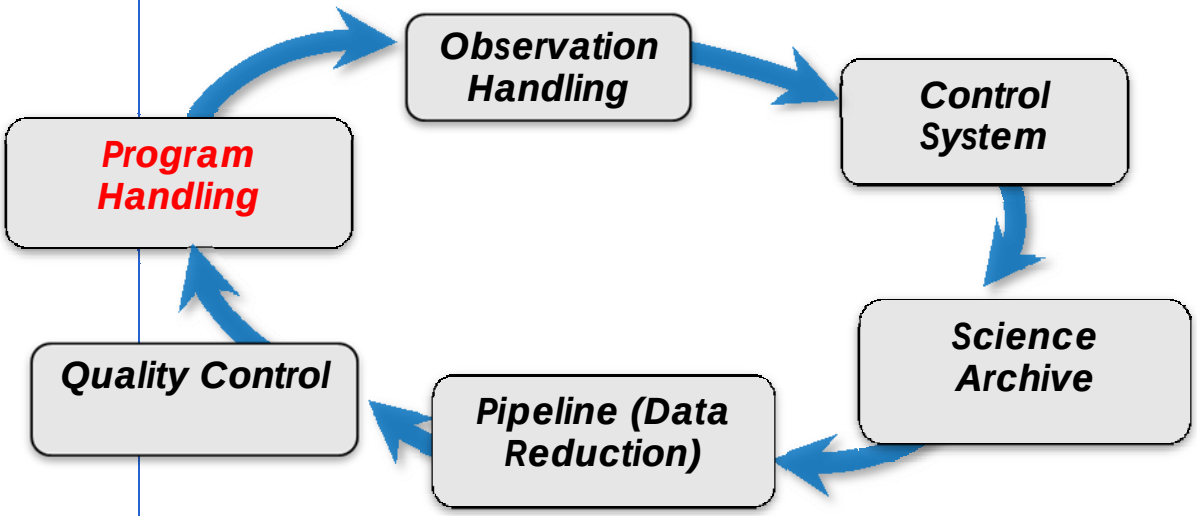
Resulting data is archived

Resulting data is processed for the purpose of quality control

Dynamic Model (OMT)



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

Program Handling (Proposal Submission)

1998

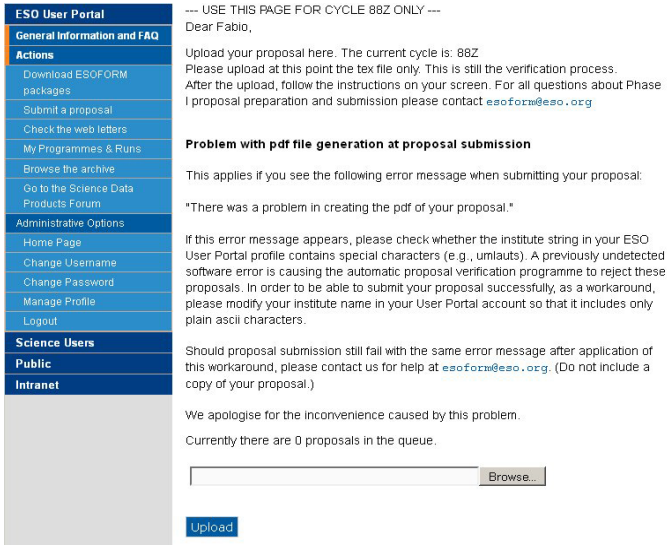
Latex packages are downloaded from an ftp server

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.

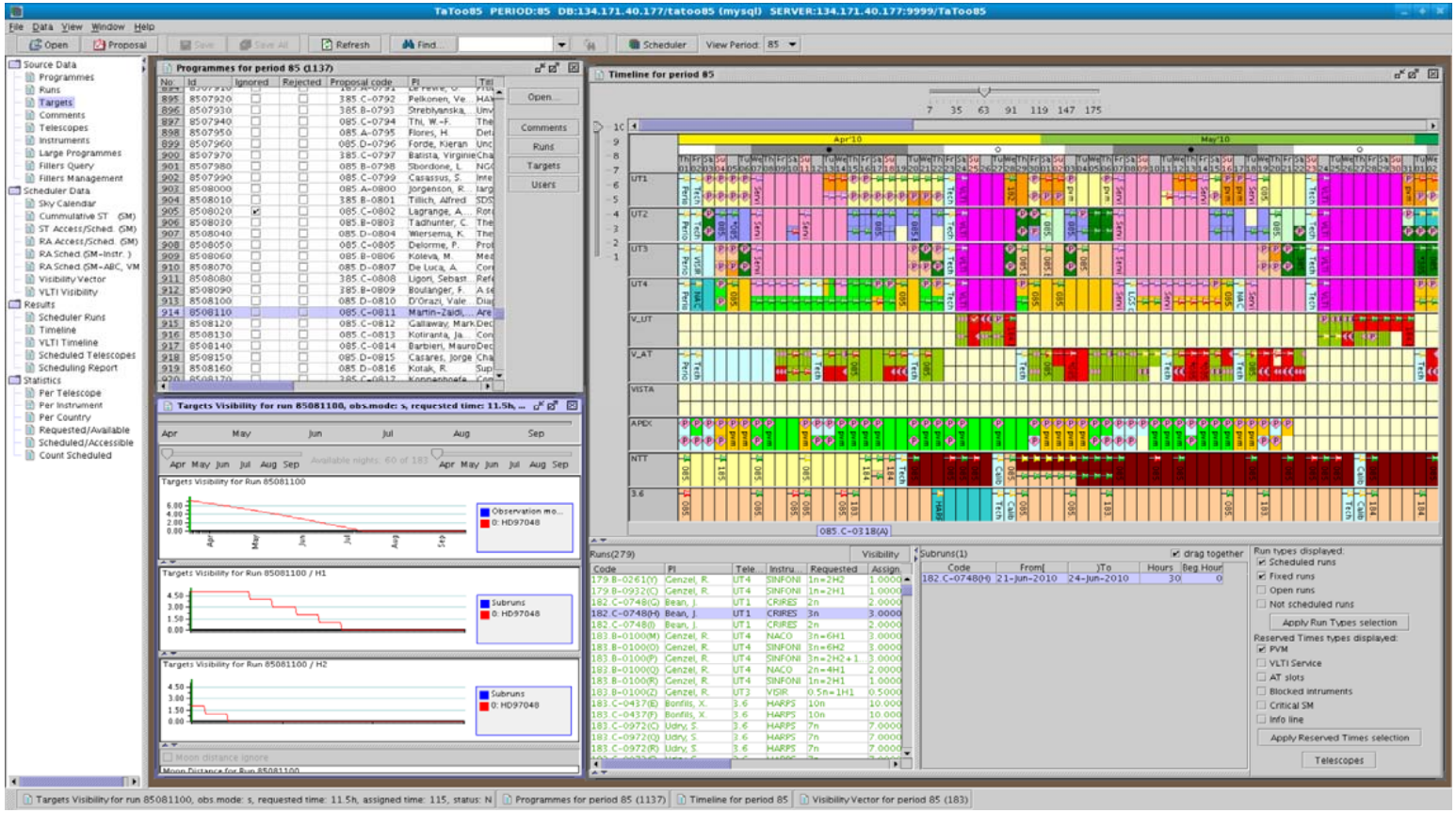
A PDF file is generated by the system and checked by users

Users submit the proposal.

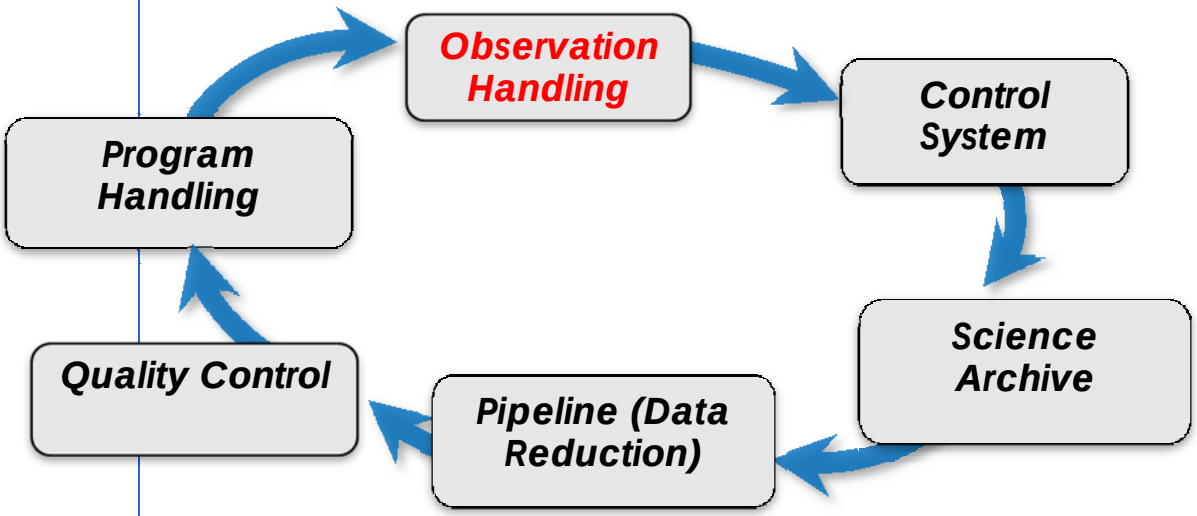
Program Handling (Long-Term Scheduling)

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



Introduction

Submission, Evaluation and Scheduling of Observing Proposals

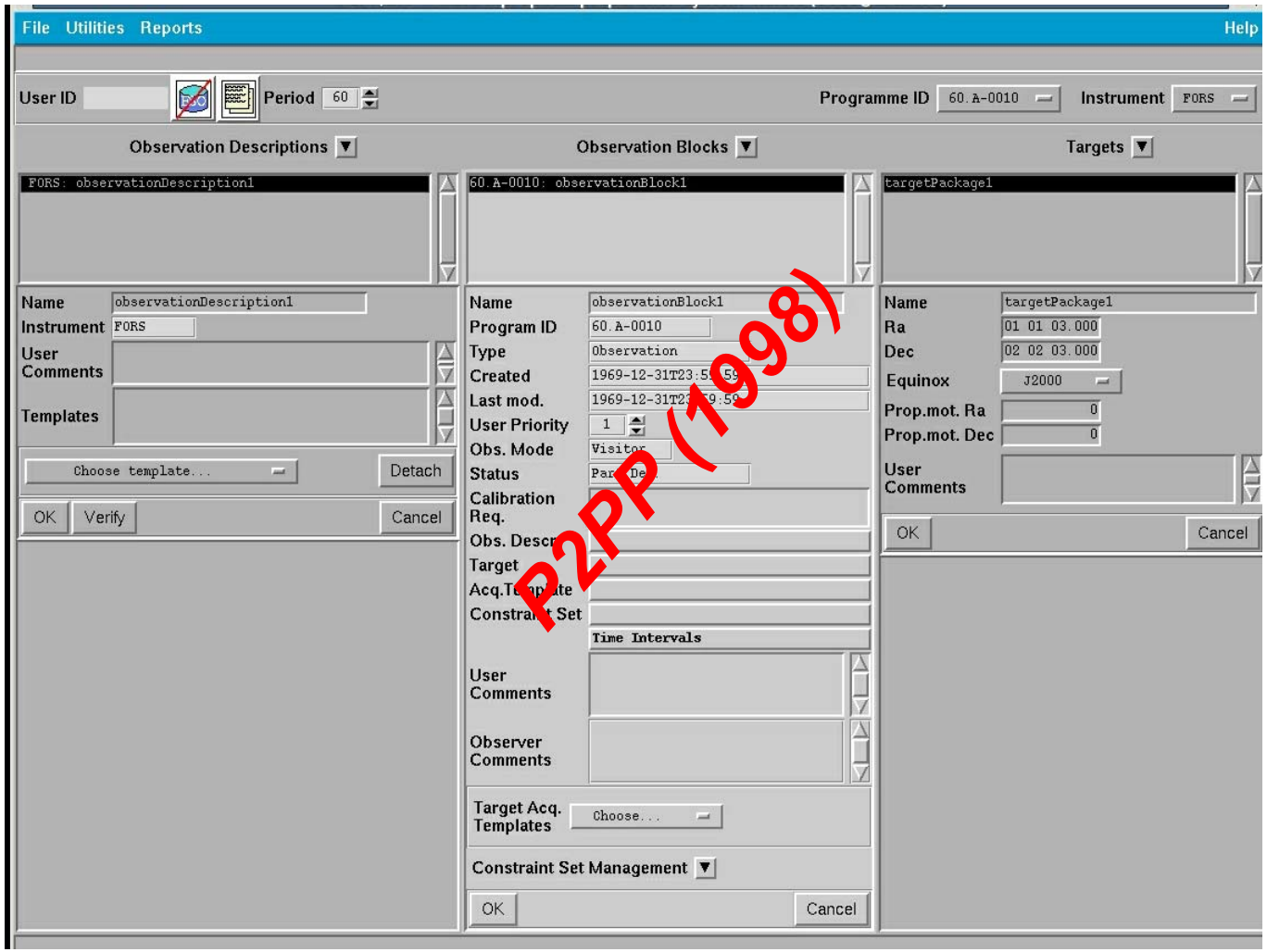
Preparation, Scheduling and Execution of Observations

Pipeline (Data Reduction)

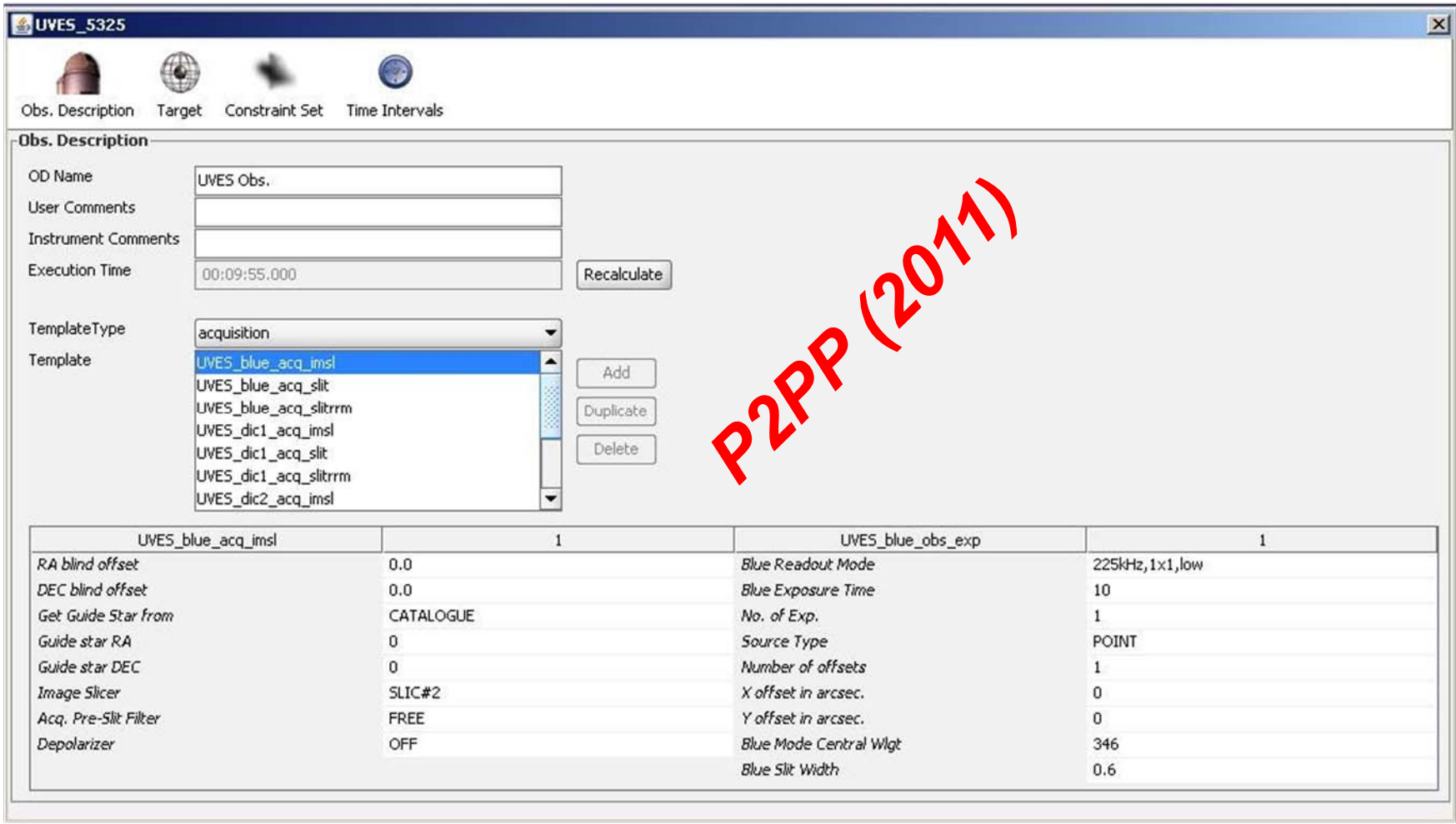
Data Archiving and Distribution

The Data Flow System for an E-ELT

Observation Handling (OB preparation)



Observation Handling (OB Preparation)

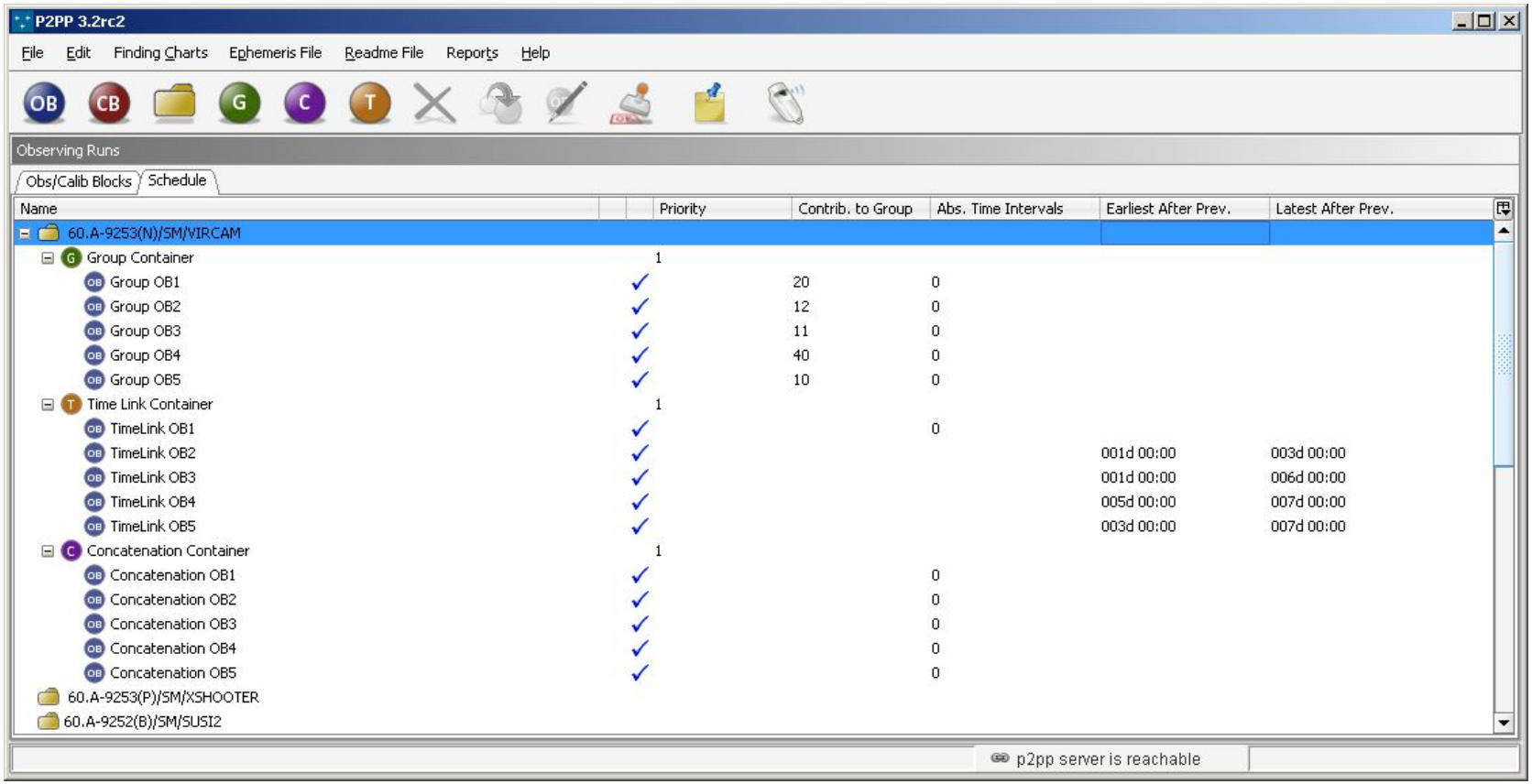


Observation Handling (OB Preparation)

Survey Telescopes (i.e., VISTA, VST) brought in new ways of observing

One program might span several years and including hundreds of OBs.

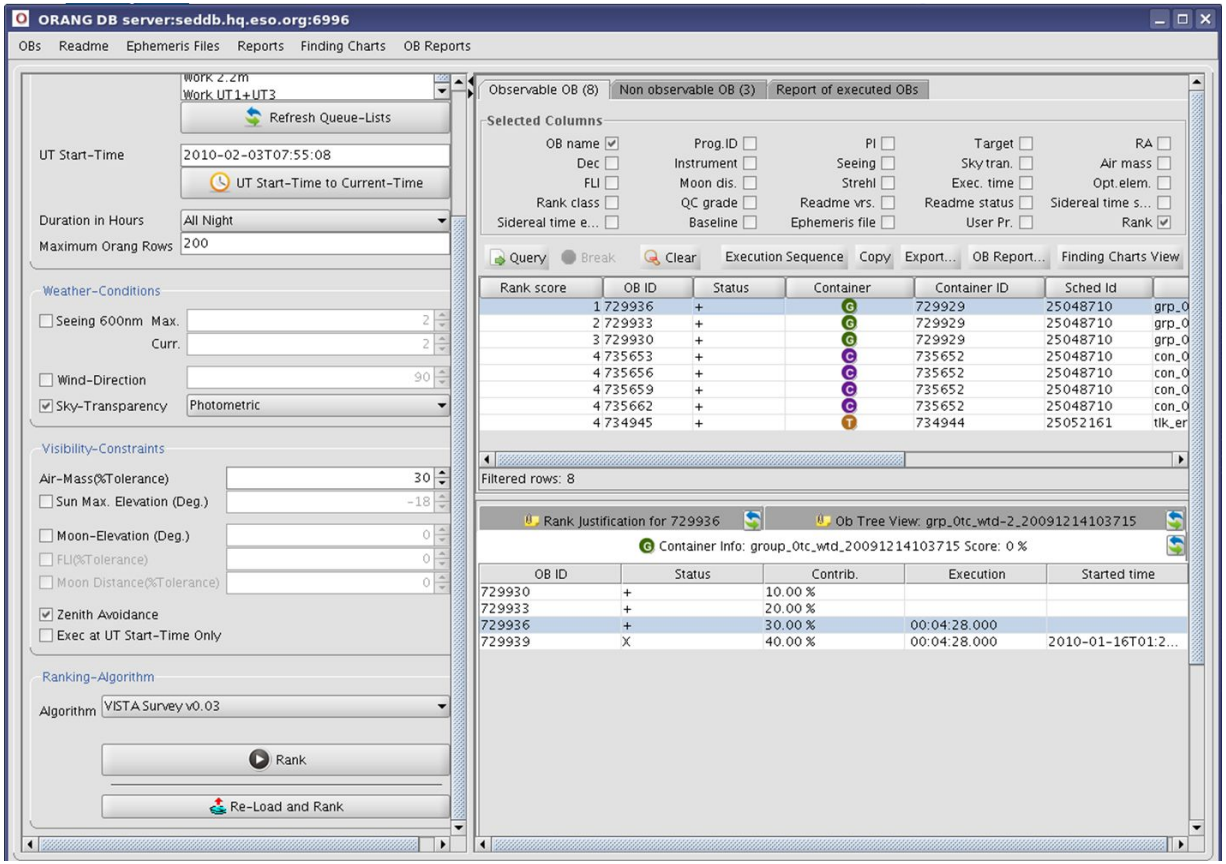
Scheduling containers allow astronomers to express more complex strategies by creating additional abstraction on top of individual OBs that allow expressing dependencies between them.



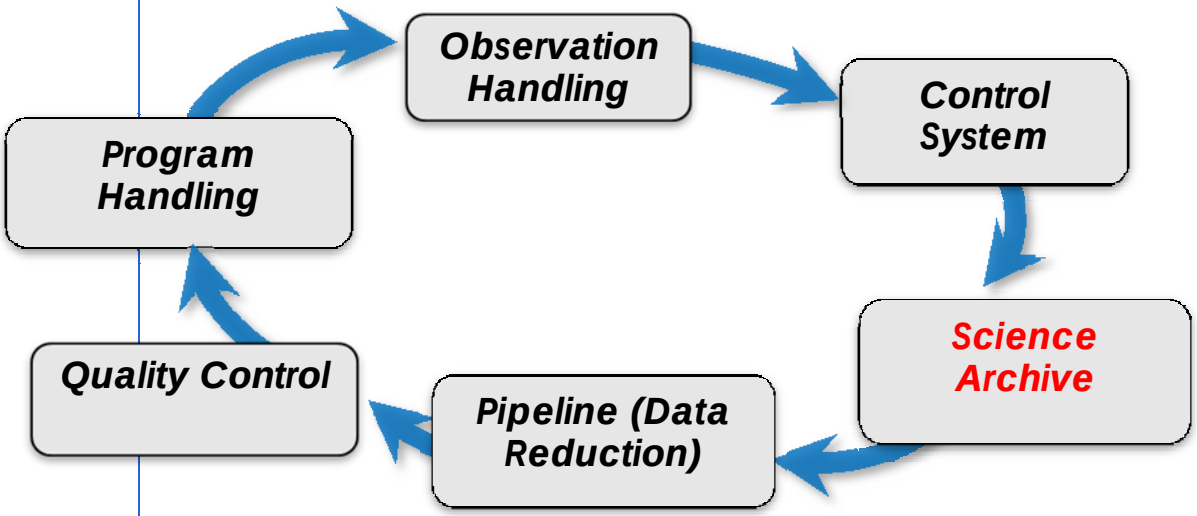
Observation Handling (OB Execution)

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

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



Science Archive



Introduction

Submission, Evaluation and Scheduling of Observing Proposals

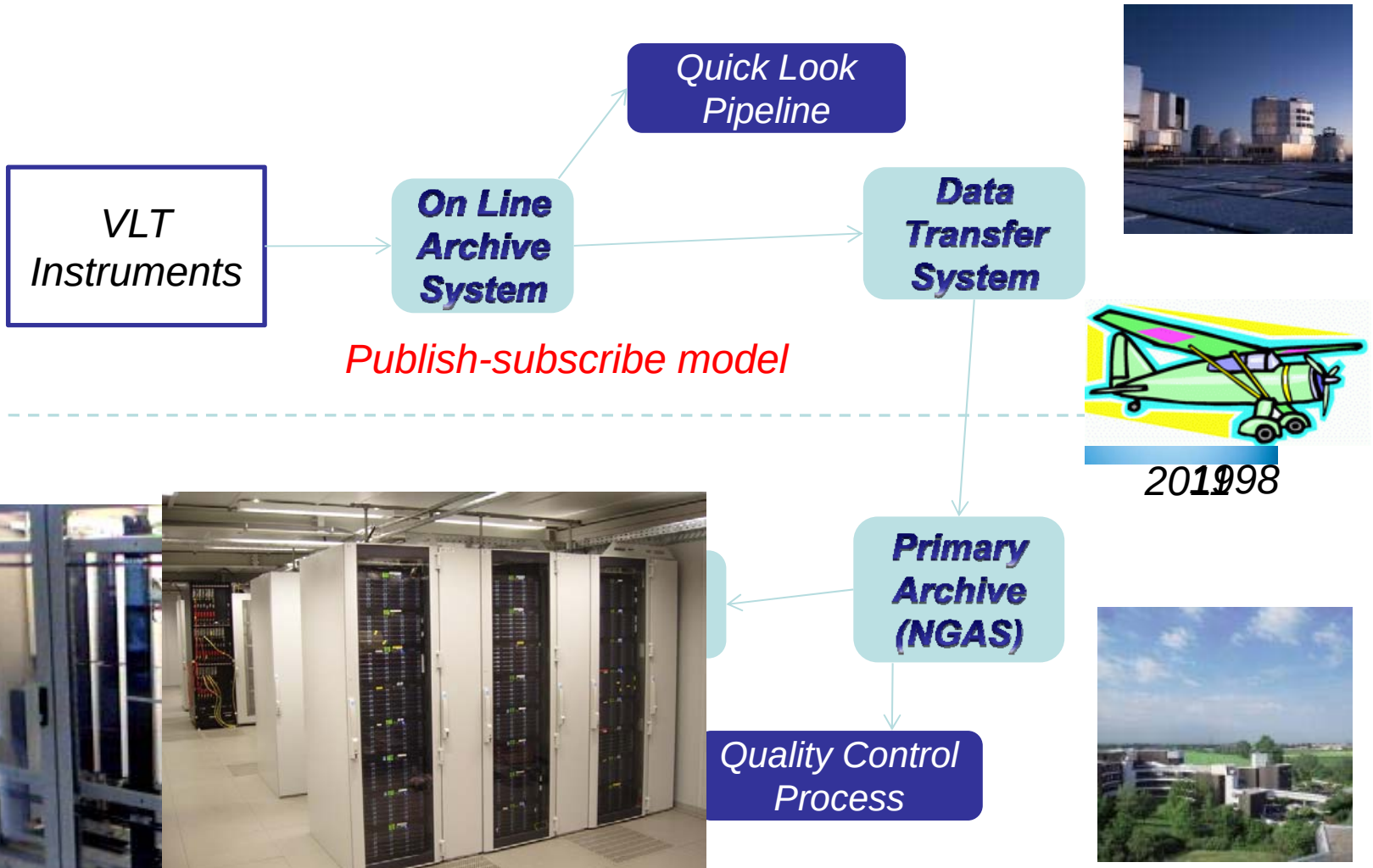
Preparation, Scheduling and Execution of Observations

Pipeline (Data Reduction)

Data Archiving and Distribution

The Data Flow System for an E-ELT

Data Flow Back End



Data Transfer and Distribution

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 Transfer Monitor

Data Flow Statistics ♦ Network Transfer Monitor ♦ Network Bandwidth Usage ♦ Network Transmission Time Stats ♦ File Search ♦ Main Page

Network File Transfer

Showing first 3026 of 3026 files with total size 22,565.4 MB

Last query: 2009-03-05 22:04:02 UTC

Define query

Creation date	2009-03-04	Date is defined to start at 12:00 UTC
fileID pattern	None	Enter one or more matching strings, separated
transfer status	all	
file type	Fits	
# of files to be shown	all	

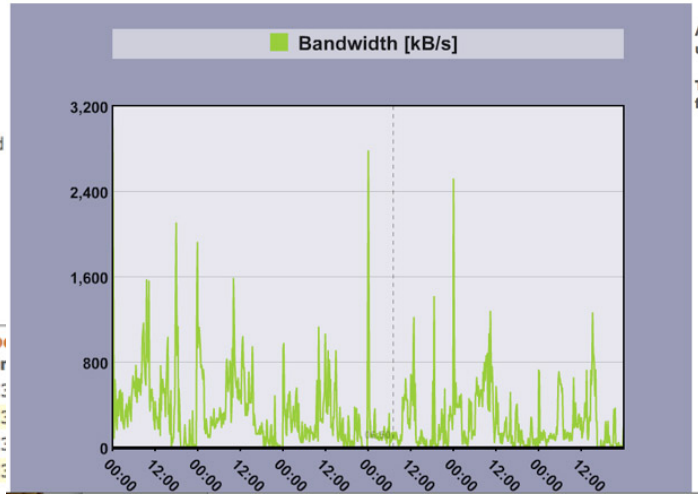
Submit

file id	status	size [MB]	priority	waiting Time [hh:mm:ss]	last up [yyyy-r
VIMOS.2009-03-05T02:07:49.834	✓	8.56	20	05:01:50	2009-03
VIMOS.2009-03-05T02:07:49.835	✓	8.51	20	05:01:51	2009-03
VIMOS.2009-03-05T02:07:49.854	✓	8.79	20	05:01:49	2009-03
VIMOS.2009-03-05T02:07:49.855	✓	8.91	20	05:01:49	2009-03

Bandwidth Usage from 2009-02-01 00:00 UTC to 2009-02-07 00:00 UTC

Show data from: 2009-02-01 00:00 (valid format: yyyy-mm-dd hh:mm)
to: 2009-02-07 00:00

Submit

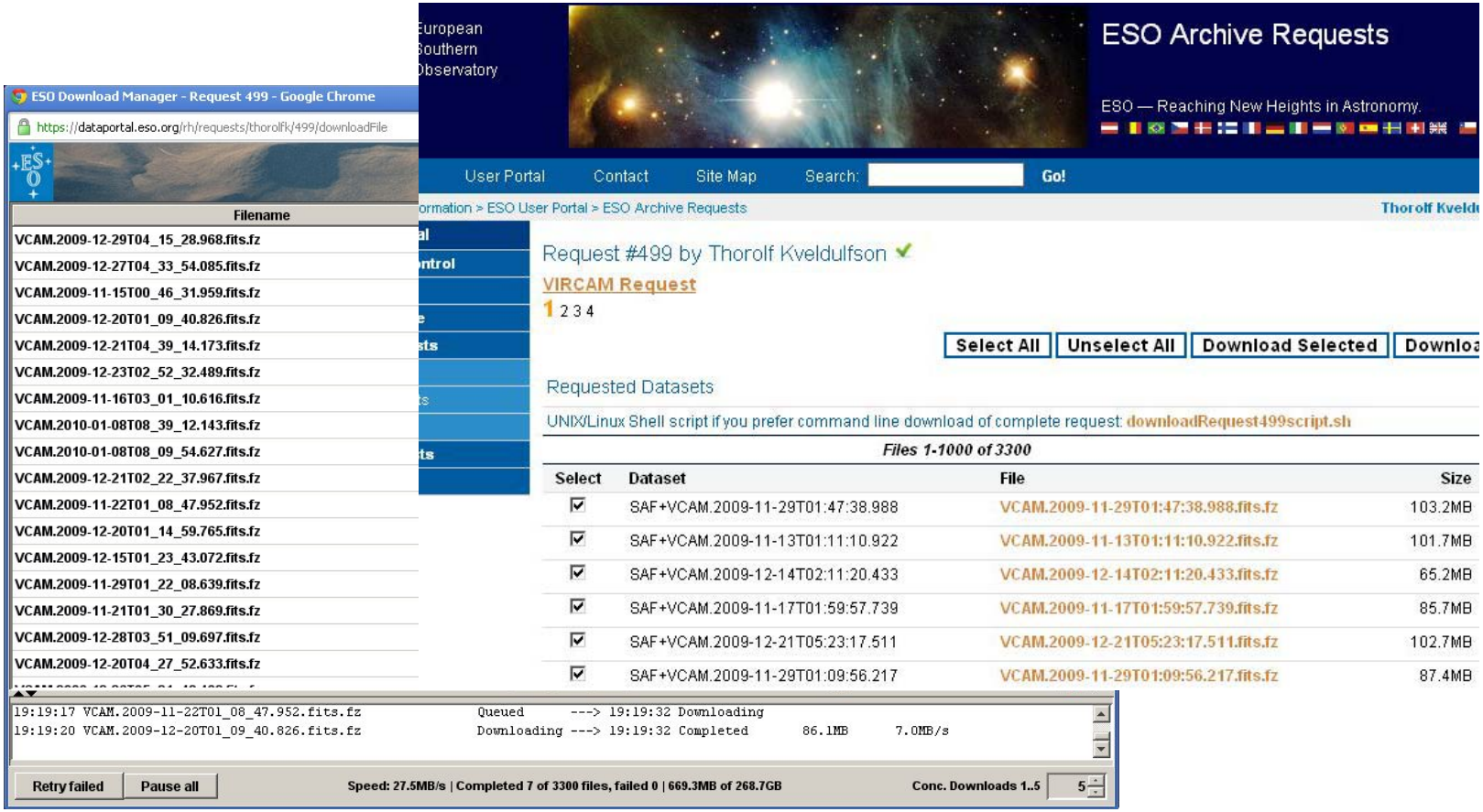


Average bandwidth usage: 293.0kB/s

Total size of transferred files: 151.9GB

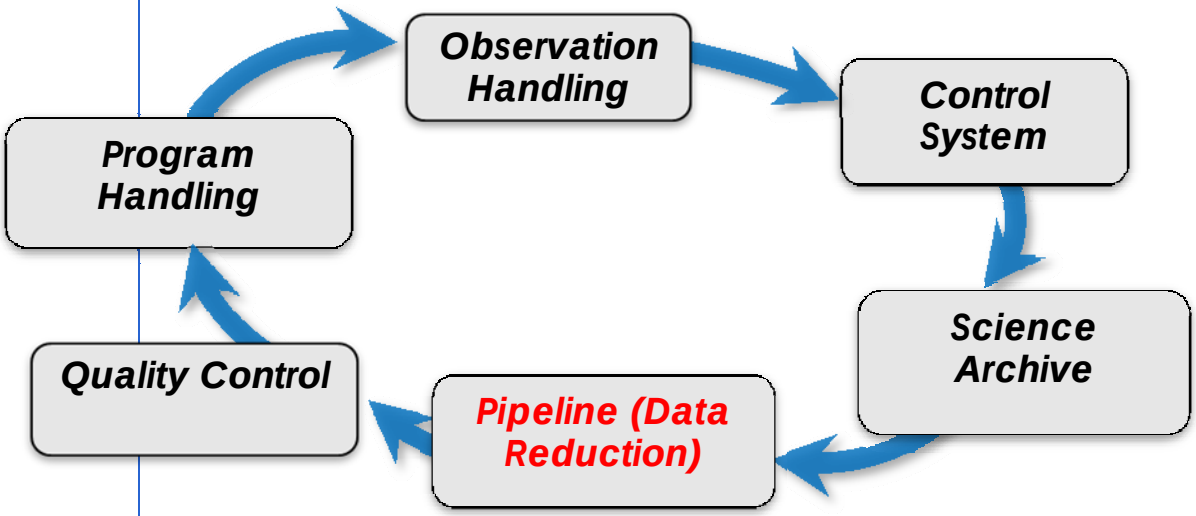
Data Distribution (request Handler)

Code re-use from CADC/ALMA



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

Pipeline (Data Reduction)



Introduction

Submission, Evaluation and Scheduling of Observing Proposals

Preparation, Scheduling and Execution of Observations

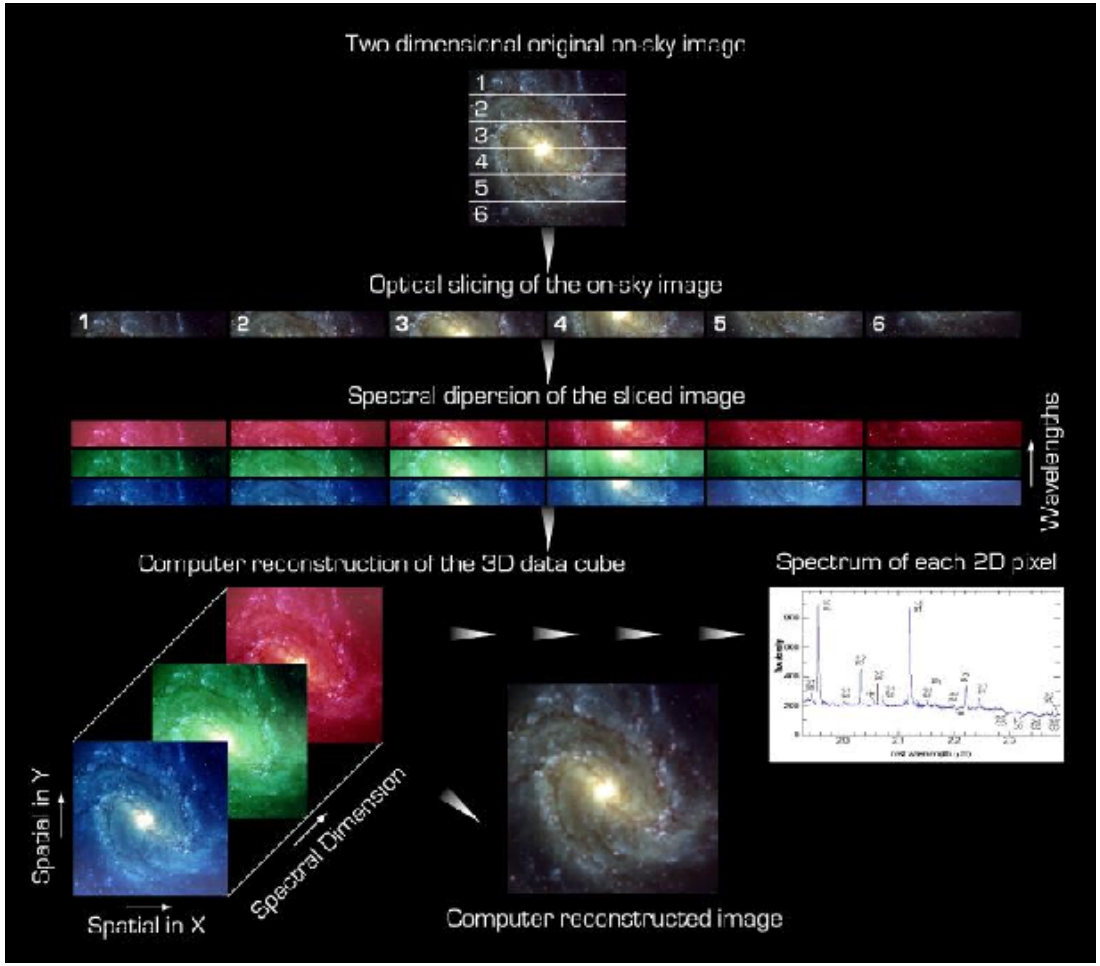
Pipeline (Data Reduction)

Data Archiving and Distribution

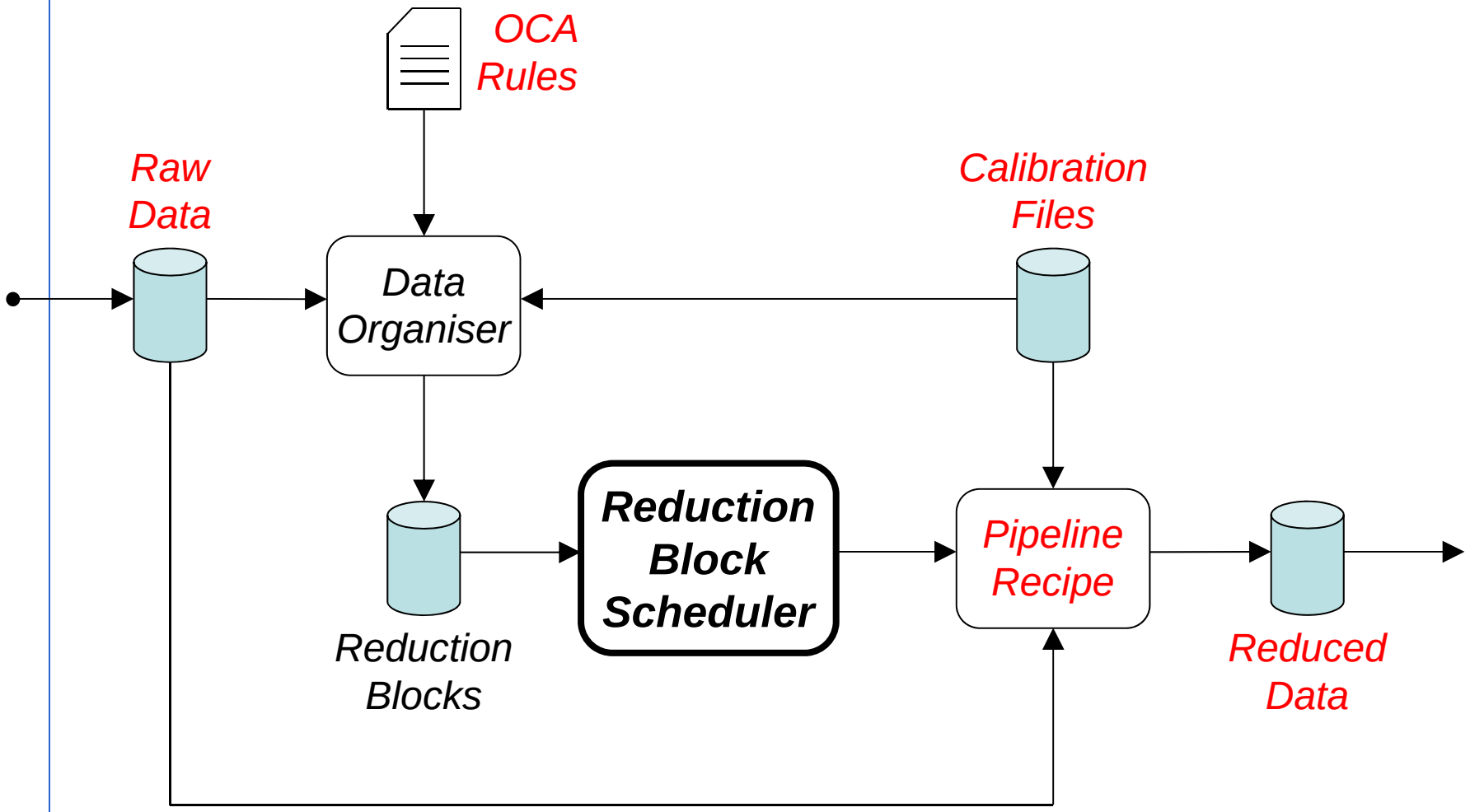
The Data Flow System for an E-ELT

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

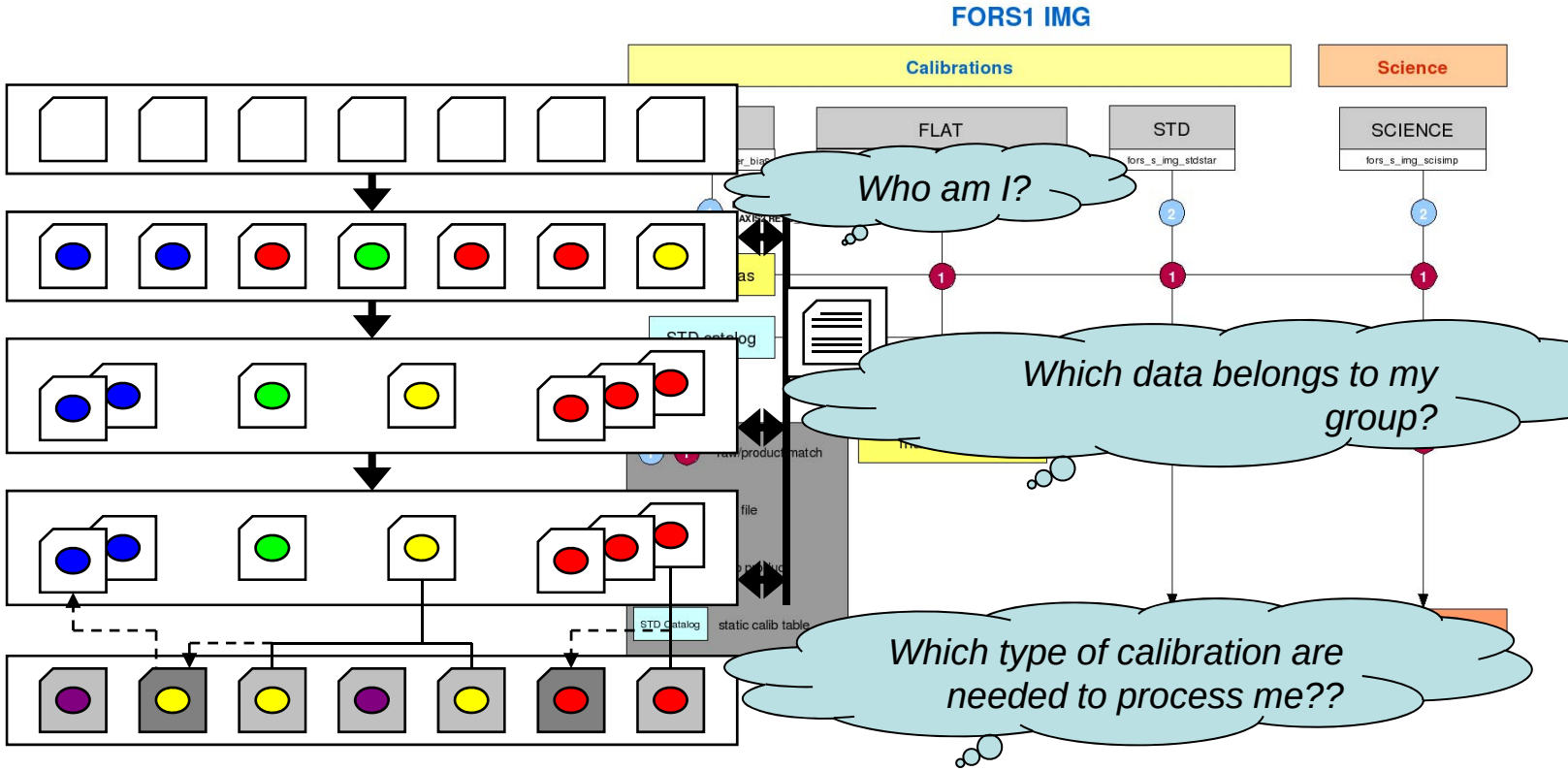


Pipeline Infrastructure



Data Reduction at the Telescope (cont)

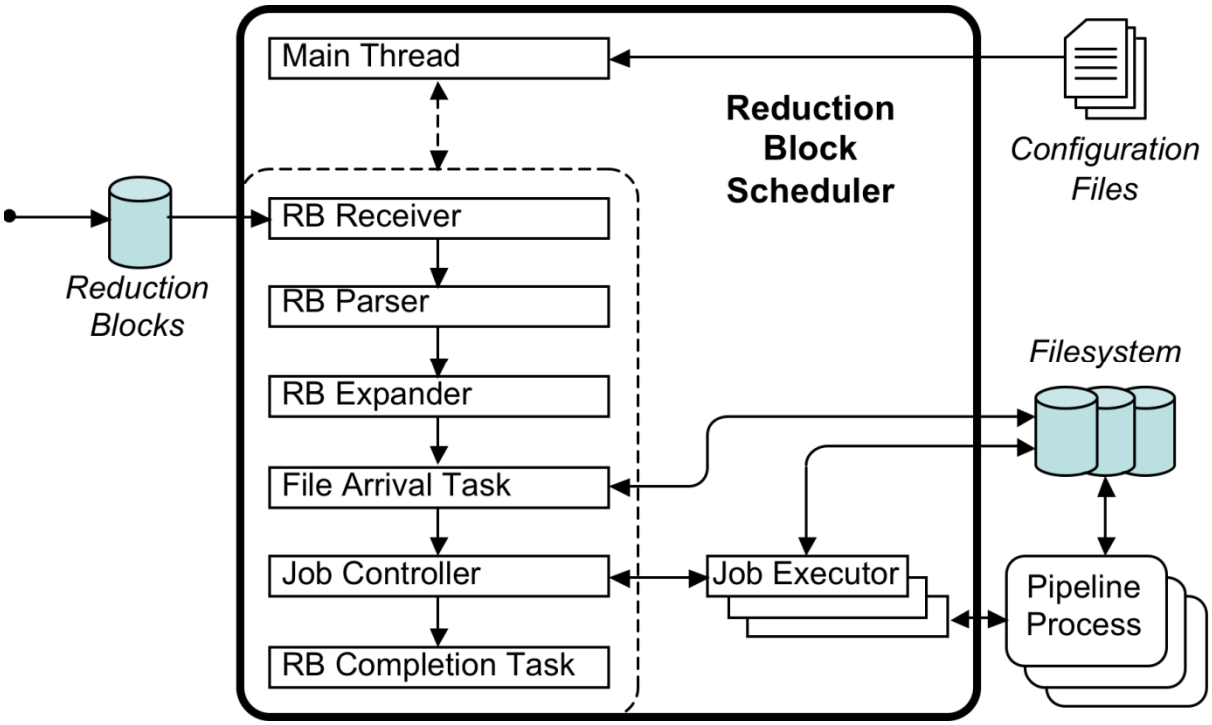
- Automatic Data Organization (available in 1998 in C++, re-engineered in Java in 2005)
- Based on a flexible rule-engine & a domain-specific language
- Creates Reduction Blocks (contains all information for reducing a set of related data)



Data Reduction at the Telescope (cont)

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

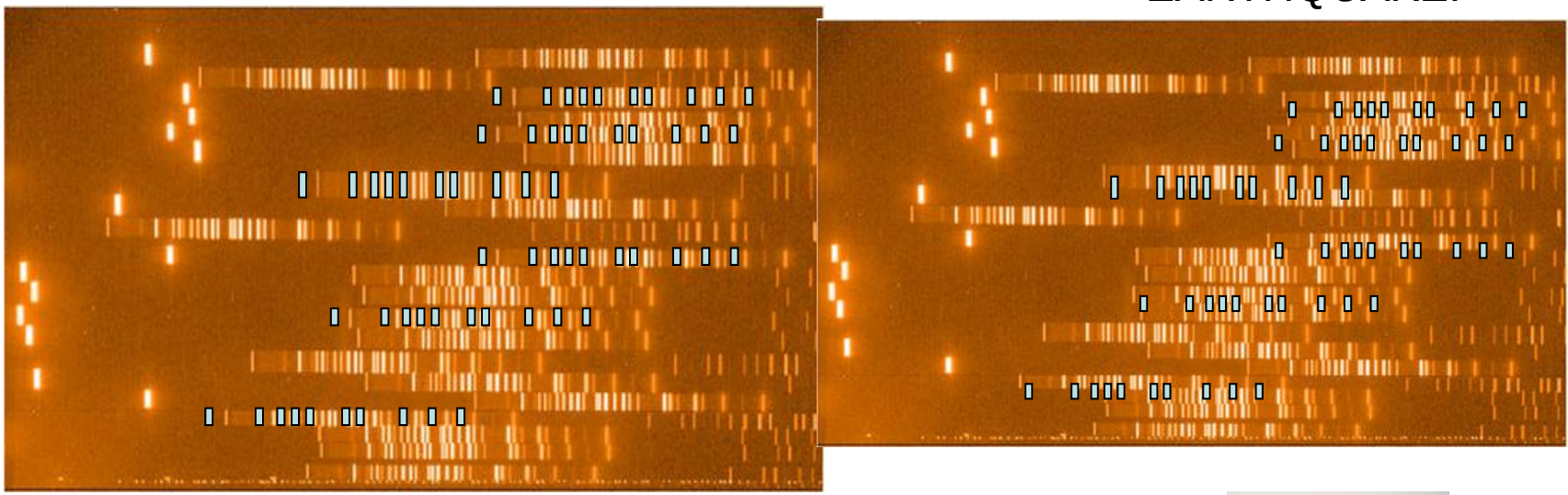
- Multi-threaded application
- Takes in account dependencies between Reduction Blocks



Pipeline Algorithms – New approaches

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

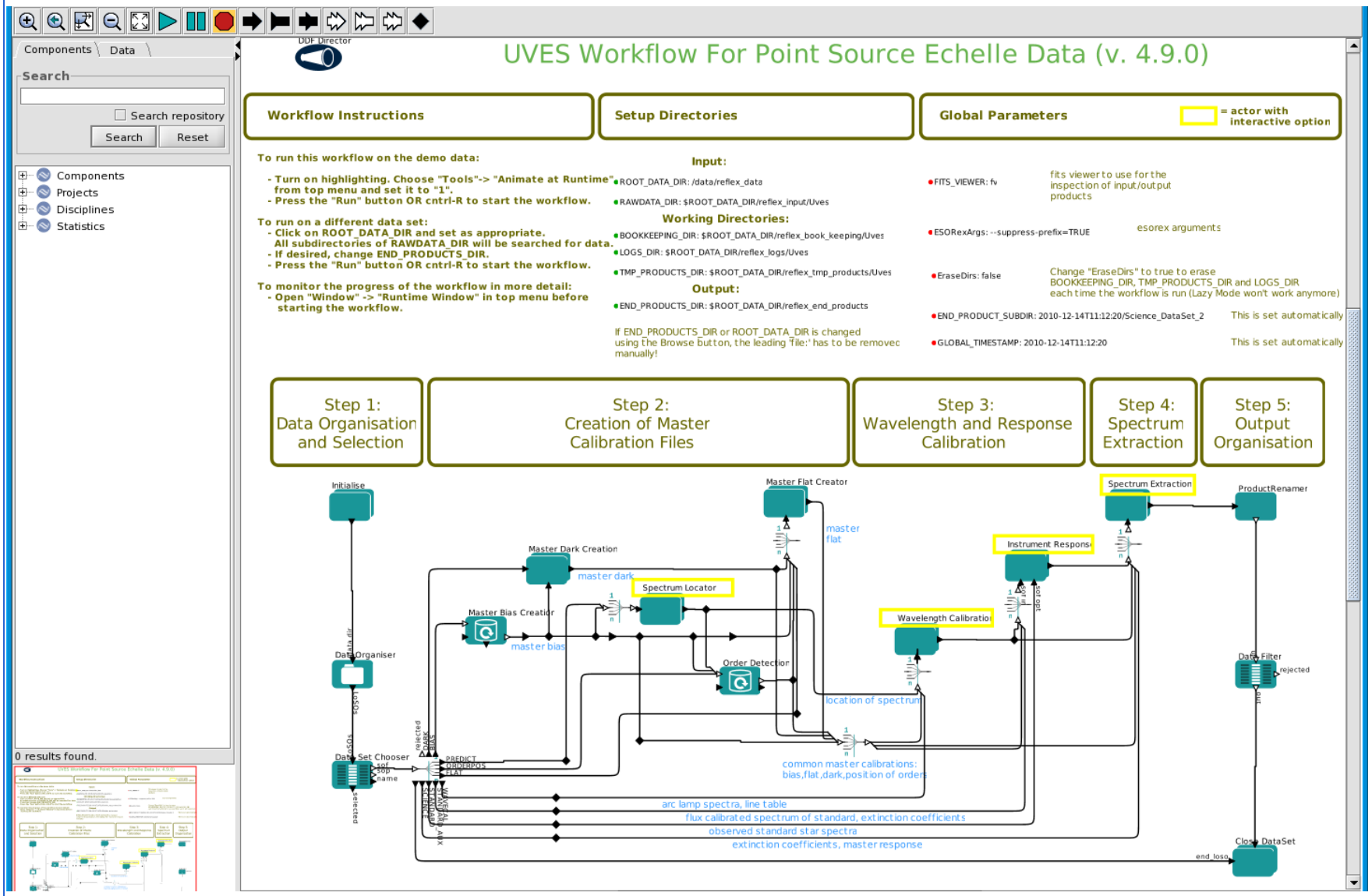
EARTHQUAKE!



New approaches (such as pattern-matching) are needed

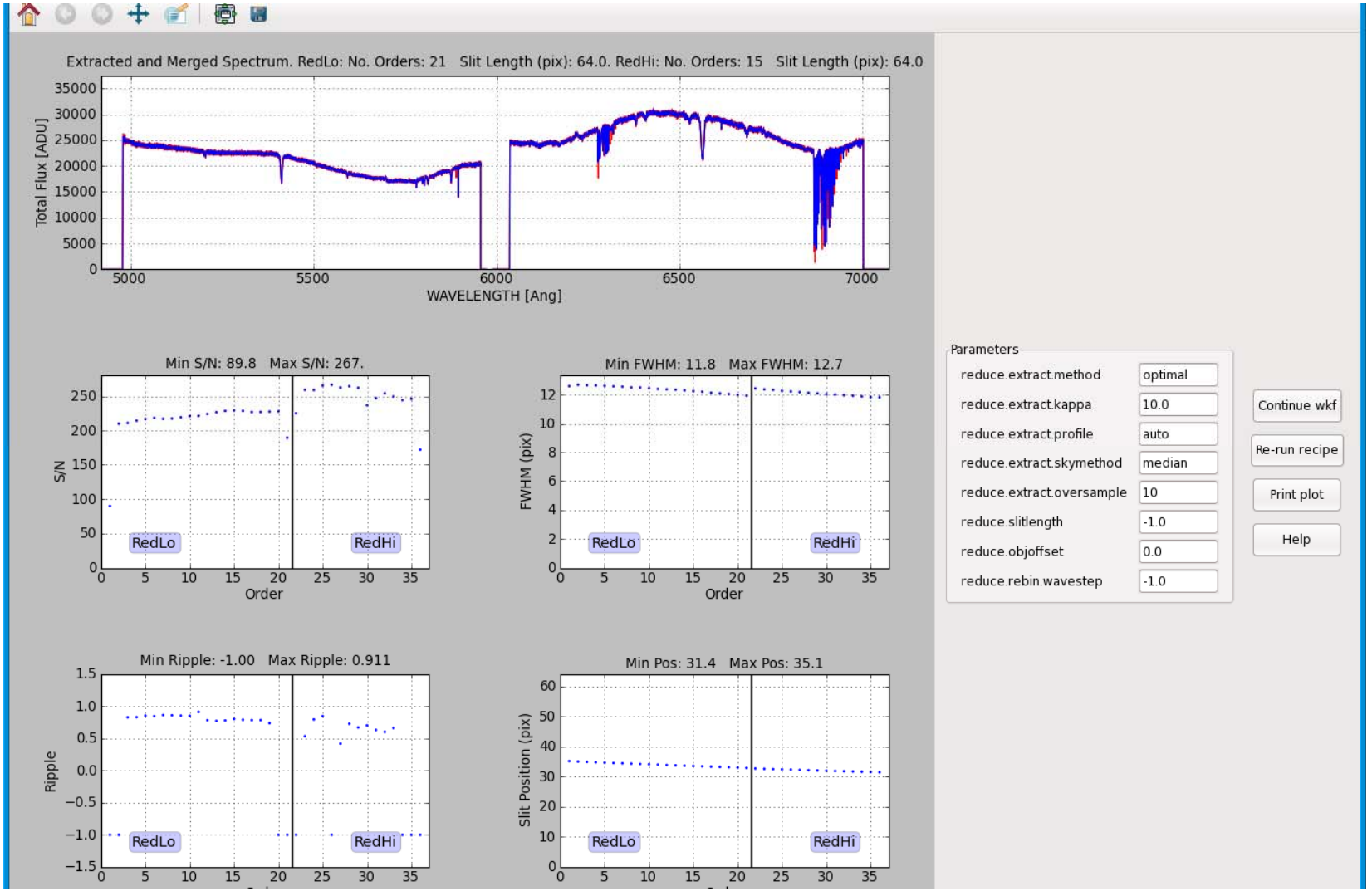
In Memoriam Carlo Izzo

Data Reduction at Home (Reflex)



Data Reduction at home (Reflex)

Demo: Ballester et al.



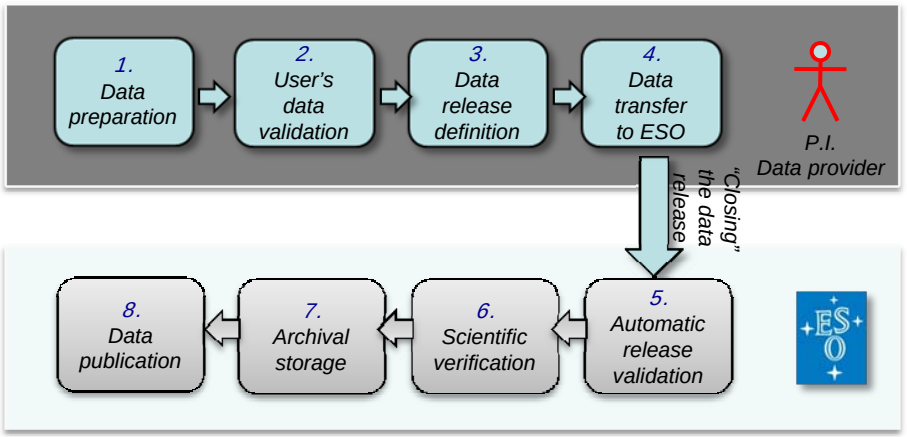
Phase 3: Handling Survey Data Products

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

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.

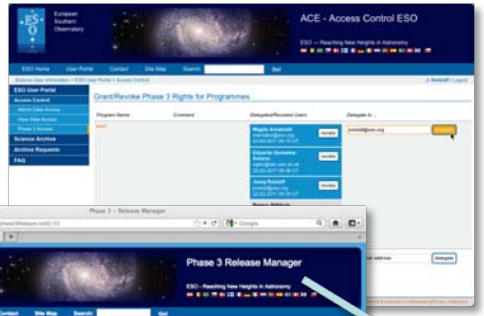
The new Phase 3 infrastructure supports the reception, validation and publication of data products from the public survey projects and large programmes to the ESO Science Archive Facility.

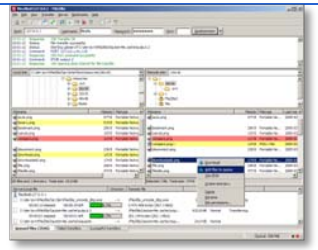
Phase 3 Process and Responsibilities



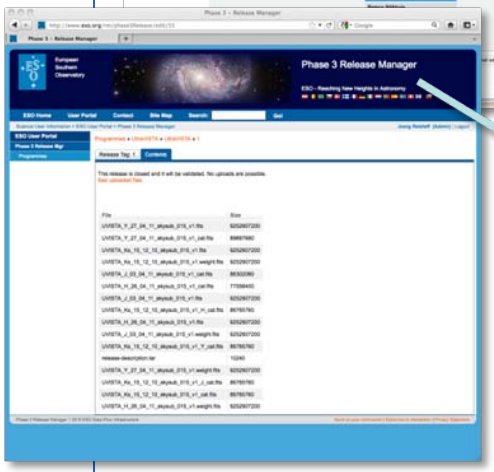
http://www.eso.org/sci/observing/phase3.html

Phase 3 data flow & infrastructure



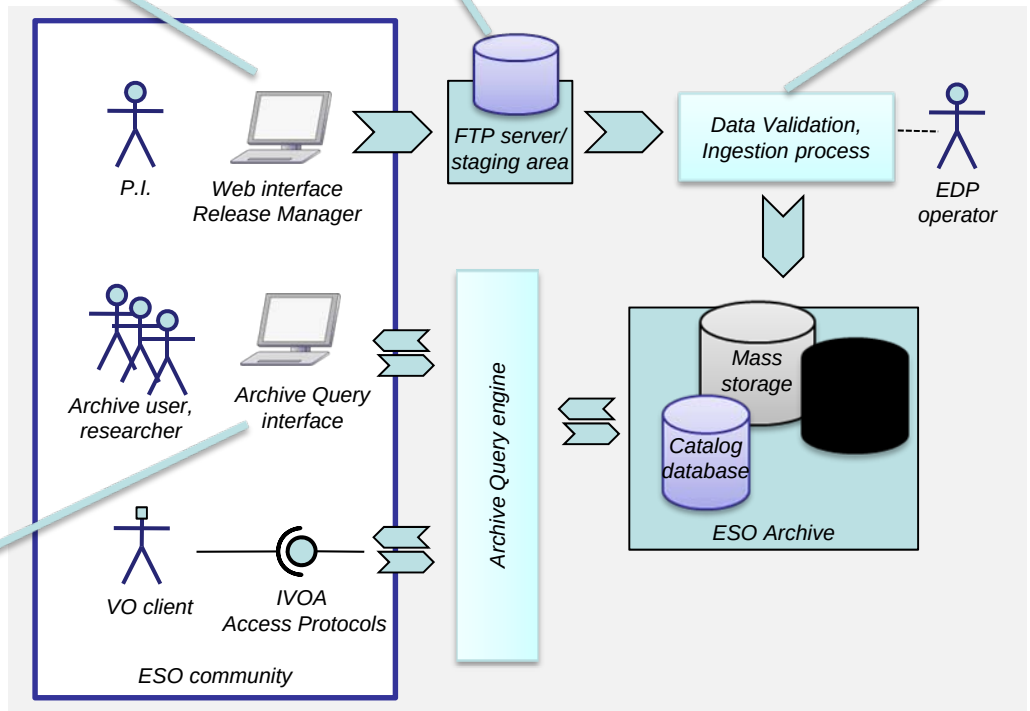


The release validator is a command-line application that helps to verify the data standard and validity of the header keywords against predefined rules.



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

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.



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

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

Start of operations: 10 March 2011

Data Flow System for the ELT

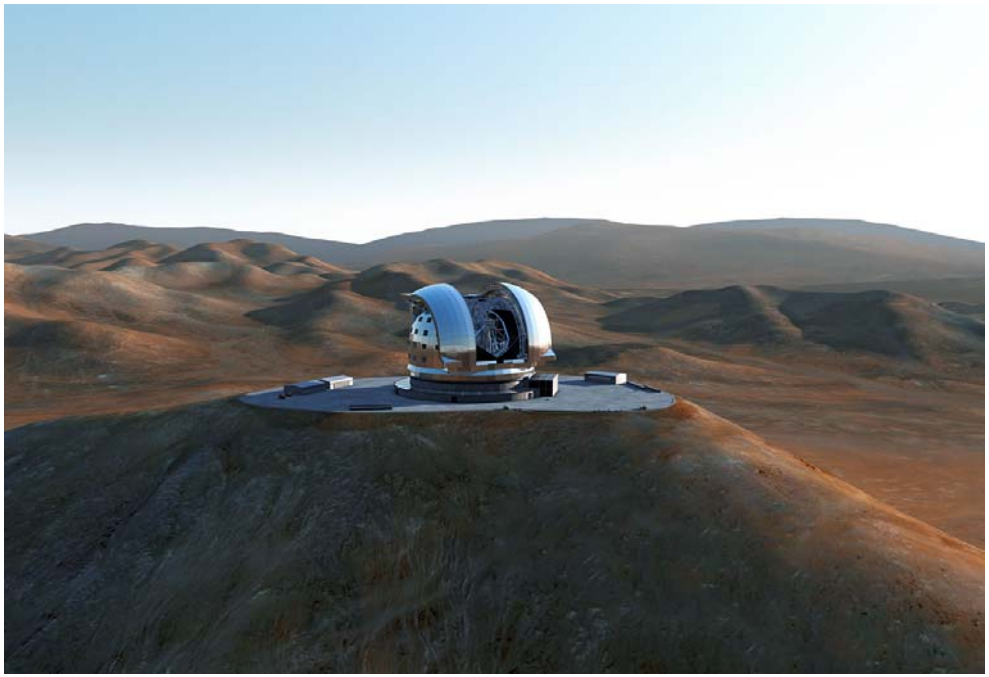
End-To-End Operational Model similar to VLT.

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