Astronomical Data Analysis Software and Systems VII
ASP Conference Series, Vol. 145, 1998
R. Albrecht, R. N. Hook and H. A. Bushouse, eds.

The VLT Science Archive System

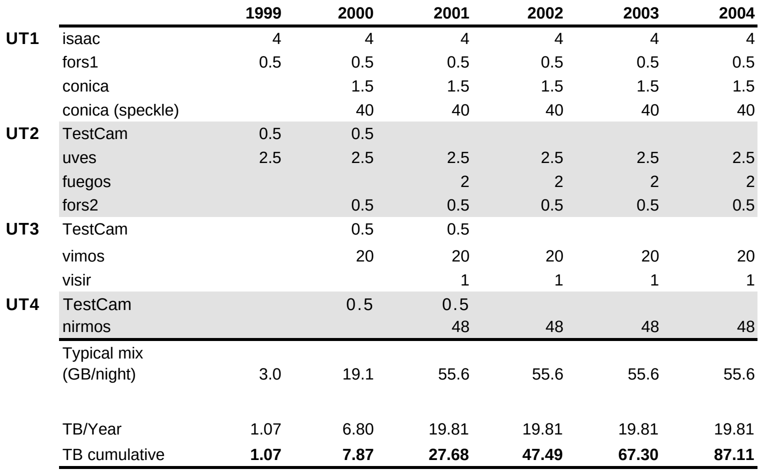
M. A. Albrecht, E. Angeloni, A. Brighton, J. Girvan, F. Sogni, A. J. Wicenec and H. Ziaeepour

European Southern Observatory, send e-mail to: `malbrech@eso.org`

Abstract. The ESO[1] Very Large Telescope (VLT) will deliver a Science Archive of astronomical observations well exceeding the 100 Terabytes mark already within its first five years of operations. ESO is undertaking the design and development of both On-Line and Off-Line Archive Facilities. This paper reviews the current planning and development state of the VLT Science Archive project.

1. Introduction

The VLT Archive System goals can be summarized as follows: i) record the history of VLT observations in the long term; ii) provide a research tool - make the Science Archive another VLT instrument; iii) help VLT operations to be predictable by providing traceability of instrument performance; iv) support observation preparation and analysis.



		1999	2000	2001	2002	2003	2004
UT1	isaac	4	4	4	4	4	4
	fors1	0.5	0.5	0.5	0.5	0.5	0.5
	conica		1.5	1.5	1.5	1.5	1.5
	conica (speckle)		40	40	40	40	40
UT2	TestCam	0.5	0.5				
	uves	2.5	2.5	2.5	2.5	2.5	2.5
	fuegos			2	2	2	2
	fors2		0.5	0.5	0.5	0.5	0.5
UT3	TestCam		0.5	0.5			
	vimos		20	20	20	20	20
	visir			1	1	1	1
UT4	TestCam		0.5	0.5			
	nirmos			48	48	48	48
	Typical mix (GB/night)	3.0	19.1	55.6	55.6	55.6	55.6
	TB/Year	1.07	6.80	19.81	19.81	19.81	19.81
	TB cumulative	**1.07**	**7.87**	**27.68**	**47.49**	**67.30**	**87.11**

The data volume expected from the different instruments over the next years is listed in Table 1. Figures are given in gigabytes for a typical night during steady state operations. Estimated total rates per night are derived by making assumptions on a mixture of instrument usage for a typical night.

[1] http://www.eso.org/

© Copyright 1998 Astronomical Society of the Pacific. All rights reserved.

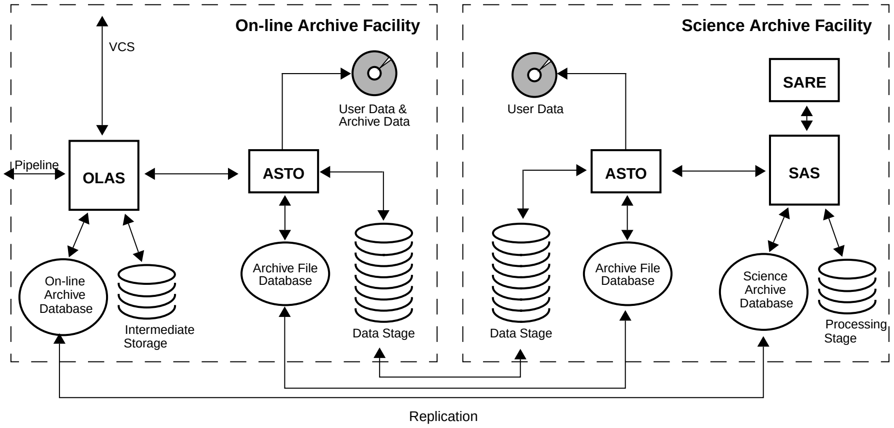


Figure 1. Overview of the VLT Archive System Architecture.

In order to achieve the goals listed above, a system is being built that will include innovative features both in the areas of technology and functionality. Among its most distinct features, the system a) will be scalable through quasi online data storage with DVD Jukeboxes and on-line storage with RAID arrays and HFS; b) will include transparent replication across sites; c) will be data mining-aware through meta-databases of extracted features and derived parameters.

2. System Architecture

The main components of the VLT Archive System are (see figure 1): the On-Line Archive Facility (OLAF) and the off-line Science Archive Facility (SAF). The On-Line Archive System (OLAS) takes care of receiving the data and creates the Observations Catalog while the Archive Storage system (ASTO) saves the data products onto safe, long-term archive media. The SAF includes a copy of ASTO used mainly for retrieval and user request handling, the Science Archive System (SAS) and the Science Archive Research Environment (SARE). The SAS stores the Observations Catalog in its Science Archive Database. All the data is described in an observations catalog which typically describes the instrument setup that was used for the exposure. Other information included in the catalog summarize ambient conditions, engineering data and the operations log entries made during the exposure. In addition to the raw science data, all calibration files will be available from the calibration database. The calibration database includes the best suitable data for calibrating an observation at any given time.

The Science Archive Research Environment (SARE) provides the infrastructure to support research programmes on archive data. Figure 2 shows an overview of the SARE setup. Archive Research Programmes are either user defined or ESO standard processing chains that are applied to the raw data. Each of the processing steps is called a Reduction Block (RB). Typically the first reduction block would be the re-calibration of data according to the standard calibration pipeline. A reduction block consist of one or more processes which are treated by the system as black boxes, i.e., without any knowledge of its im-

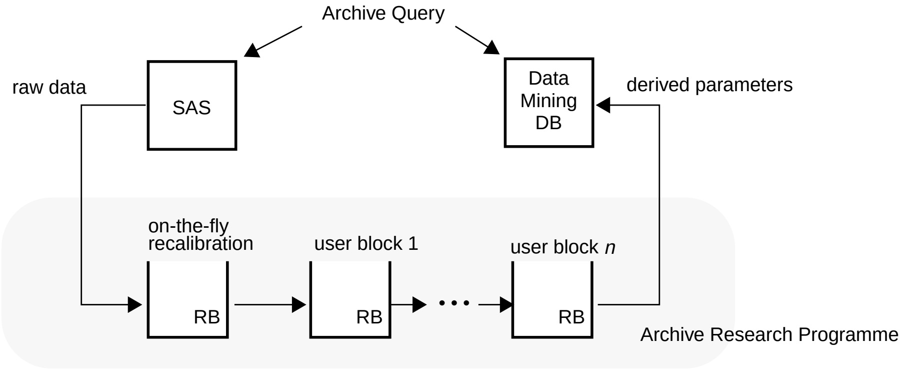


Figure 2. Overview of the VLT Science Archive Research Environment.

plementation. However, the reduction block interface (input and output data) do comply to a well defined specification. This feature allows any reduction module to become part of the chain. In fact, this flexible architecture also allows the research programme to analyze different kinds of data from images and spectra to catalogs and tables of physical quantities. The output of an archive research programme will be derived parameters that are fed into the data mining database.

3. Data Mining in the Science Archive Research Environment

Observation data will be stored within the VLT Science Archive Facility and will be available to Science Archive Research programmes one year after the observation was made.

However, in face of the very large data amounts, the selection of data for a particular archive research project becomes quickly an unmanageable task. This is due to the fact that even though the observations catalog gives a precise description of the conditions under which the observation was made, it doesn't tell anything about the scientific contents of the data. Hence, archive researchers have to first do a pre-selection of the possibly interesting data sets on the basis of the catalog, then assess each observation by possibly looking at it (preview) and/or by running some automated task to determine its suitability. Such procedure is currently used for archive research with the HST Science Archive and is acceptable when the data volume is limited (e.g., 270 GB of WFPC2 science data within the last 3.5 years of HST operations).

Already after the first year of UT1 operations, the VLT will be delivering data quantities that make it not feasible to follow the same procedure for archive research. New tools and data management facilities are required. The ESO/CDS Data Mining Project aims at closing the gap and develop methods and techniques that will allow a thorough exploitation of the VLT Science Archive.

One approach at tackling this problem is to extract parameters from the raw data that can be easily correlated with other information. The main idea

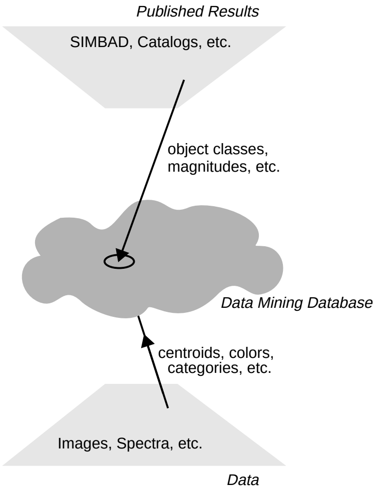


Figure 3. Overview of the data mining environment.

here is to create an environment that contains both extracted parametric information from the data plus references to existing databases and catalogs. In its own way, this environment then establishes a link between the raw data and the published knowledge with the immediate result of having the possibility to derive classification and other statistical samples. Figure 3 illustrates the general concept.

An example of a semi-automatic parameter extraction is the object detection pipeline used by the ESO Imaging Survey (EIS) Project. Every image in the survey is subject of a set of reduction steps that aim at extracting object parameters such as 2-D Gaussian fitted centroids, integrated magnitudes, etc. The cross-correlation of parameters of this kind with selected databases and catalogs (e.g., eccentric centroids with galaxy catalogs) would provide a powerful tool for a number of science support activities from proposal preparation to archive research.

4. Conclusions

The VLT Archive System being developed will provide the infrastructure needed to offer the Science Archive as an additional instrument of the VLT. The main capabilities of the system will be a) handling of very large data volume, b) routine computer aided feature extraction from raw data, c) data mining environment on both data and extracted parameters and d) an Archive Research Programme to support user defined projects.