TITLE
ASP Conference Series, Vol. **VOLUME**, **PUBLICATION YEAR**
EDITORS

Towards a Framework for Ground-based Optical/Infrared Public Surveys

Luiz da Costa

European Southern Observatory, Karl-Schwarzschild-Str 2, Garching bei München, D-85728, Germany

Abstract.

In this paper the requirements for establishing a framework for conducting ground-based, optical/infrared public surveys are discussed. The main challenges facing the implementation of this type of project are outlined as well as the bearing they have on the design of a survey system. The experience accumulated over the past three years by the ESO Imaging Survey (EIS) project at ESO has served to set requirements for the new end-to-end survey system being developed. The survey pipeline software has been designed to meet the needs of long-term, multi-instrument public imaging surveys carried out by a small team. Some of its main features are also briefly described.

1. Introduction

The start of regular operations of several 8-m class telescopes has created a new demand for large imaging surveys. However, in contrast to other ongoing all-sky surveys, there is a pressing need for surveys capable of providing data sets from which samples reaching magnitudes comparable to the spectroscopic limit of the large-aperture telescopes can be drawn. Fortunately, relatively large area, deep surveys have become possible thanks to the development of wide-field imagers and 2-m and 4-m class telescopes being used for large imaging programs.

The need for imaging data is especially acute for the ESO community at large, considering VLT's four telescopes and the wide array of spectrographs available, some lacking imaging capabilities. Foreseeing this need ESO established in 1997 the ESO Imaging Survey (EIS) project in charge of conducting public imaging surveys in support of VLT science. In the course of this program, many important lessons have been learned ranging from organizational matters to a vast array of technical issues which must be addressed in order to develop a proper framework for conducting such programs over long stretches of time. Sharing this experience may be valuable for similar programs envisioned in the near future such as the VST and VISTA projects.

In section 2 an attempt is made to differentiate between various possible types of public surveys, their terms of reference and their impact on the design of a survey system. In section 3, EIS is used as a case study to illustrate the impact that the goals and the specific constraints of a particular survey have on the design of a survey pipeline software. Some of the main features of the most recent version of the EIS survey pipeline are also briefly discussed. Finally, in section 4 a brief summary is presented. More information on the EIS project can be found at the URL www.eso.org/science/eis.

2. General Considerations

The need for imaging surveys is of course not new - what has dramatically changed in the past few years is the rapid increase in the rate digital images are being collected and consequently the volume of data to be processed. The challenge is now to build efficient systems to store, process, access, visualize, analyze and scientifically exploit the resulting large data sets. The basic reduction steps are well-known and include: removal of instrument signature; astrometric and photometric calibration; source detection and measurement of flux and structural parameters; and object classification (star/galaxy). Less standard steps include the combination of information obtained using different passbands and/or telescopes, stacking and mosaicing of images, preparation of color catalogs for faint sources and use of multi-wavelength information for object classification. Depending on the duration of the program it is desirable to consolidate these various steps into a pipeline enabling the un-supervised reduction of the data.

For long-term programs a greater degree of autonomy and complexity may be required to ensure efficiency, uniformity and consistency in the preparation of final products. The pipeline requirements may differ depending on the nature (public/private); the size of the survey program and the terms of reference. There is now a wide spread recognition of the value of planned public imaging surveys, designed to meet needs of a broad community (*e.g.* ESO, NOAO, VISTA). However, the framework within which these efforts should be conducted is still a matter of debate and varies from place to place. While many ideas are in common, such as the need for an oversight committee and community engagement, they differ in the amount of telescope time allocated to these programs, the composition and role of the survey team, the scope of the survey pipeline and in the procedures and policies regarding the distribution of reduced data and products.

It is important to emphasize that not all "pipelines" are or should be alike. In fact, the design of a survey pipeline depends on various factors including: the specific goals of the survey and its design (number of passbands, area, depth and type of fields); expected input data rate (frequency of observations, survey strategy) and total data volume (expected duration of the program); the number of telescope/detector configurations involved; the type and variety of desired products and how frequently these products should be made available. One must also consider the desired level of uniformity of the data, photometric accuracy and seeing tolerance which will in turn set requirements for data quality control. All of these points translate into requirements in the design of the survey software. Even though the points listed above may sound obvious, there is a general tendency to underestimate the impact they have on the pipeline design and its development timescale.

To be successful a public survey depends not only on the pipeline design but also on the context in which it is inserted. It must rely on the infrastructure available and on a variety of interfaces with other areas of activities. It is crucial to reach an equitable balance between the specific responsibilities of the survey team and the institute at large. Some important points to consider are: whether the telescope used is a survey-dedicated or a general-purpose one; the nature of the telescope/survey team interface; the availability of tools to prepare and monitor observations; the availability of calibration data and how efficiently raw data can be accessed by the survey team as well as their overall uniformity.

At the back-end of the pipeline one must also consider whether tools and resources are available to distribute data and define the level of on-line resources that can be supported. The latter is particularly important as the currently available technology precludes the wide spread distribution of pixel maps. The alternative is to provide more advanced products, in the form of source catalogs (single passband, color, mosaic), while allowing users to retrieve sections of images to assess by themselves the quality of the products for their scientific goals. Of course, this adds a new layer of complexity to the survey pipeline. Furthermore, in contrast to surveys carried out from space and radio surveys, ground-based optical/infrared surveys are sensitive to weather and atmospheric conditions. This implies that, in addition to derived products, information regarding the seeing conditions, the photometric quality of the nights and the completeness of the survey must also be provided both periodically and at the time of release of a given set of data. This is particularly true when releasing incomplete data sets, as may be the case if a pre-established timetable of data releases is adopted. Adequate tools to automatically monitor the progress of the survey and a suitable interface to the web pages are required for regular updates based on information stored in the pipeline database. In addition, at the time of data releases this information should be properly summarized in supporting publications for reference.

Arguably an overriding concern in establishing a public survey is the definition of the size, composition and responsibilities of a survey team. Of course this depends on the specific plans and objectives of each institute, the availability of resources and the particular circumstances of each public survey. It also depends on the lead time available for planning, for carrying out the software development and the complexity of the particular implementation of the survey pipeline. Needless to say, building a standardized, well-tested pipeline takes time and it is only possible by systematically dealing with real data. Furthermore, some level of development and maintenance will be required over the lifetime of the project to improve the performance of the system, extend its scope and adapt it to new observing strategies and technologies. On the other hand, the permanent involvement of a science team to monitor the progress of the observations and to validate the images and derived products is indispensable, regardless of how automatic the pipeline has been designed to be. Therefore, a combination of scientists and scientific programmers will always be required. The size of the team will depend on the phase of the project. It is important to point out that for a successful recruitment the program must be attractive from both the scientific and technological point of view. This raises the important issue of what the scientific return for the people involved in public surveys should be. Unless one finds an acceptable solution to this problem recruiting will be a difficult task.

Careful consideration of the many issues mentioned above is essential to establish the requirements for survey pipeline software and the general framework within which a long-term, sustainable public survey can be carried out. Satisfactorily addressing these issues is critical to ensure the timely release of the data in a form that enables the entire community to make the best use of the public survey data.

3. The EIS Experience

The ESO Imaging Survey project was established by ESO to provide data sets of interest to VLT. It was started about one year before the commissioning of the first VLT unit telescope (UT1). In its first phase the program relied exclusively on the NTT, carrying out both a wide-angle survey to moderate depth and a deep optical/infrared survey of the HDF-S and AXAF fields. Besides the short-term goal of finding potential targets for the period of science verification of UT1 and the first two periods of regular operation, the project was also seen as an important test case to: establish a framework for conducting public surveys at ESO; drive the software development for the implementation of a survey system; create mechanisms to tap on the expertise spread throughout the ESO community; disseminate know-how throughout the community by training students and post-docs and by making available survey software and tools; and create a testing ground for software development using real data.

To this end the first action was to organize a Working Group assembled from scientists, representing a broad range of scientific interests, drawn from the community and in charge of: designing and monitoring the progress of the surveys; preparing proposals and progress reports to the time allocation committee; suggesting to the ESO management policies regarding data distribution and timetables; serving as representatives of the scientific goals of the community.

The Survey Team was assembled via a special Visitor Program which funds short and long-term visits of known experts in the field as well as students and post-docs. The responsibilities of the team include: 1) preparation and execution of observations; 2) data reduction; 3) production and verification of calibrated images, object catalogs and targets; 4) maintenance of supporting database; 5) maintenance of web pages; 6) periodic reports describing the progress of the observations and publications describing the data and the contents of each data release. In the particular case of EIS the observations, reductions and the development of the survey pipeline were carried out in parallel.

EIS started as short-term experimental project and in its early phase a pragmatic bottom-up approach was adopted. The main effort was to assemble and interface existing software to enable raw images to be processed and astrometrically calibrated, mosaics created and object catalogs produced with a minimum of supervision for a specific instrument and survey strategy. However, many of the administrative tasks, photometric calibration, data verification and archiving remained external to the pipeline proper and required human intervention. While this first implementation of the pipeline was effective for the input data rate and the total volume of survey data expected at the time, this approach fell short of the requirements of an end-to-end system. The short-comings of this first implementation became evident when different telescope/detectors were used (*e.g.* infrared; multi-chip), the data volume greatly increased due to the use of a

$8K \times 8K$ wide-field imager and the status of the program changed from a short experimental program to a long-term one.

Using different telescope/detector combinations and survey strategies entailed handling: optical and infrared data which require different reduction procedures; detectors with different number of ports and chips; different FITS images and headers; different combination of optics, CCDs and filters. Altogether EIS observations have been carried out at two telescopes using four different instruments and a variety of filters, for a total of 156 nights in 36 runs over a period of 3 years. The large data volume required the installation of DLT jukeboxes for storing images at different stages of the pipeline and procedures to efficiently store/retrieve these data. Finally, the long-term status of the program combined with the large turnover of visitors (over 25 in three years, averaging 5-6/year) highlighted the need for a fully integrated, end-to-end, user-friendly pipeline.

Based on this previous experience over the past year the EIS team has concentrated its efforts on building a significantly expanded survey system. The main goal is to achieve stable and reliable operation, capable of: preparing the observations based on pre-defined priorities; monitoring the progress and overall performance of the survey; reporting the results of each run; sustaining a large throughput of data; dealing with different sky surveying modes and telescope/detector configurations; combining data from different passbands and telescopes; incrementally releasing fully calibrated images; verifying the quality of the products (images and derived source catalogs); identifying astronomical objects of interest and ensuring the reproducibility of the results.

Some of the major issues being addressed are: a comprehensive supporting database; a search engine for easy access to information and data, based on different combination of queries; versioning of all the software, configuration files and products; integration of all tools to manipulate and verify the data into a common environment; logging, timing and error reporting of all processes; a user-friendly pipeline and database interface.

Though ambitious, the new design is essential in order to ensure that the survey work as currently envisioned by ESO is sustainable in the long-term with a small survey team. While the public surveys conducted by ESO have been so far relatively modest in size, with VST and possibly VISTA the trend is for them to increase in size. The present framework provides a solid base from which one can ramp-up to cope with the much larger data rates expected from these telescopes in a routine and consistent way.

Arguably the most important and time-consuming step has been the design, development and ongoing integration of a supporting database. To facilitate its design, implementation and maintenance, a specific graphic interface to SQL has been developed in Python. The database contains all the information characterizing different telescope/detector configurations (*e.g.* optical throughput, CCD quantum-efficiency, filter transmission) and surveys (pointings, filters, region, integration time, number of exposures). This information is essential to ensure the consistency of the data products, for supervising the preparation of the observations and enable the survey team to keep track of the survey status. The idea is to produce directly from the pipeline database observation blocks as required by ESO's VLT Data Flow System (DFS). The DFS provides the necessary tools to keep track of the completed observations in such a way that this information can be ingested back into the database greatly facilitating the scheduling of new observations, the overall control of the survey status and the flow of updated information to the users via an interface between the pipeline and the web. In addition, the database stores identifiers and information about all raw images processed by the pipeline as well as any subsequent intermediate or final product (images or catalogs), tapes and their contents. Therefore, any product can be easily retrieved via the search engine.

To address the rapid turn-around of team members, with varying degree of experience, all the pipeline processes have been wrapped by Python scripts which are launched through a graphic interface. This interface consists of three-levels. At the top level there are control panels which encompass all the processes for the production of a given final product. One of the panels drives the reduction of raw images all the way to the production of single passband catalogs on a run basis; the second drives the preparation of final images by stacking or mosaicing pre-reduced images on a survey/passband basis; the third drives the production of either single passband or color catalogs. While default configurations exist for the survey mode, all the pipeline processes are easily configurable. The second level of the hierarchy allows the user to execute step by step the various processes wrapped together in the first-level. This is useful for testing and training as well as for setting up suitable configuration files for the batch mode. At this level there are a large set of algorithms and visualization tools which are used for the scientific validation of the data using simple statistics as number counts, color-magnitude diagrams, color-color plots, PSF monitoring, angular correlation function and some basic data-mining tools which are used for both verification of color catalogs and target selection. Finally, at the lowest level there are the individual tools used to handle images and catalogs in a variety of ways.

4. Summary

By now the need for public surveys is generally recognized both as a scientific tool in its own right and as the means to produce an astronomical database for the increasing number of large-aperture telescopes. Perhaps not yet fully appreciated is the magnitude of the effort required to produce an end-to-end survey system satisfying all the requirements for a stable and reliable operation of high-throughput, running un-supervised for long stretches of time and making data and information available to the community in a timely manner. Building a proper framework for carrying out such programs is essential to meet the expectations of the community and win their continuing support. The main purpose of this paper has been to use the experience gained by the ongoing EIS project to lay out some of the issues that need to be addressed by pipeline designers to evaluate the complexity and timescale required for the development of customized survey systems.

Acknowledgments

I would like to thank current and former members of the EIS team for their enormous effort over the years.