The Sampo Project

As an in-kind contribution to its joining fee, the Republic of Finland is funding a three-year project to assess future data reduction and analysis software needs for the ESO community. The Sampo project, named after a mysterious artifact and source of power in the Finnish Kalevala legends, will gather requirements, assess technologies and perform detailed pilot software projects to help prepare ESO for the software challenges of data handling in the next decade.

Richard Hook[1], Palle Møller[1], Janne Ignatius[2], Kari Vasko[2], Michele Peron[1], Peter Quinn[1]

[1]ESO
[2]CSC, The Finnish IT Center for Science, Finland

The data analysis systems used by astronomers in Europe, and elsewhere, were originally developed in the 1980s, or even earlier. For analysis of optical and infrared data sets, the most popular software systems are IRAF, ESO-MIDAS and IDL whilst AIPS is still being used for interferometry. Although those systems have evolved well to cope with changing hardware and new operating systems, the fundamental underlying infrastructure is unchanged. Each of them is monolithic and provides all the components needed by both users and software developers: data access mechanisms, a scripting language, a parameter interface, graphics and display packages.

In the past few years many observatories, including ESO, have placed emphasis on the development of instrument-specific procedures (which ESO calls "recipes") to remove the instrumental signature and create calibrated data products in physical units. It must be possible to run those recipes in an automatic manner (in which case they are referred to as pipelines) as one of their primary goals is to produce the quantitative information necessary to monitor instrument performance. Whilst the first ESO instrument pipelines were developed as ESO-MIDAS applications, it rapidly became apparent that ESO-MIDAS was not well suited for automatic use. In addition, ESO-MIDAS pipelines cannot deal with the huge increase of data volumes from newer VLT instruments as they do not integrate well with frameworks providing parallel processing facilities on clusters of computers. The decision was made to move away from ESO-MIDAS and instead develop pipeline recipes as C modules which can be called from different environments. Recipes may be invoked either from an automatic pipeline running at the observatory or on the desktop through a Graphical User Interface tool (Gasgano) or a command line application (Eso-Rex), which allow users to reprocess their data offline. These tools do not, however, provide the facilities for detailed analysis of calibrated data. Data complexity is not always well addressed by standard packages either. As an example, the visualization of 3D spectroscopic data requires very specific applications that have been built outside any existing environment. As a consequence, users are currently forced to install a suite of heterogeneous tools and change between different environments several times if they wish to reprocess and analyse raw data generated by complex instruments.

The ESO community has made it clear that better support in these areas should be a high priority. Although it is apparent that a new approach will be needed, it is still unclear what exactly will be required. Do we need a "new IRAF" or a "new ESO-MIDAS"? This is unlikely to be the correct approach, as such a system would have to become a monster if it were able to tackle data complexity and data volume and it would inevitably suffer from many of the defects of the older systems. In addition, millions of lines of code have been written as applications within the standard packages. We cannot just ignore them and start from scratch. We also generally prefer to avoid self-contained instrument packages, supporting all the steps from data reduction to data analysis, as this approach may lead to a great deal of redundancy and create major maintenance difficulties. It seems we need an environment in the broader sense of a variety of software elements (libraries, tools, standards, applications) that work together, or separately to allow efficient data handling. We need some way of ensuring access to legacy algorithms and experience. And we must be ready for the time, in the not too distant future, when data will not be necessarily located on a local disc. The rapidly evolving virtual observatory (VO) efforts are defining standards and new data access methods. A clear indication of the sorts of facilities that will be available is given by the recent Euro-VO *(www.euro-vo.org)* demonstrations. Over the next few years the VO will make the transition from numerous rapidly developing pilot projects into an operational reality. Any new software system will need to facilitate convenient access to the VO, both as a data producer and consumer. The requirements of this new world of data are a primary driver for future developments.

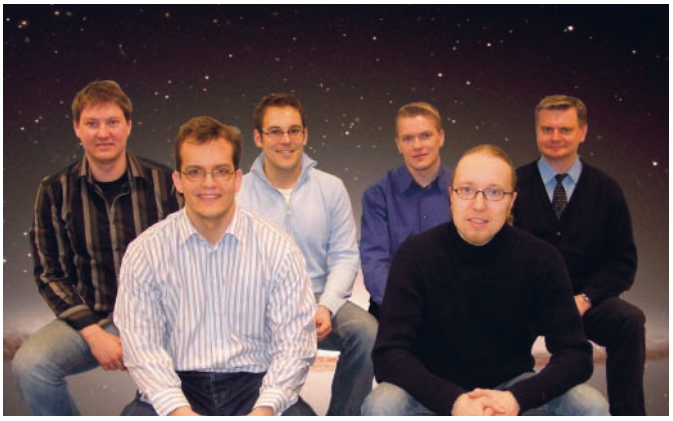
Background to the Sampo Project

Before the summer of 2004, when Finland joined ESO, the organization only had resources for developing VLT, VST and VISTA instrument pipelines. However, the ESO long-range plan recognized the need to address the question of a data analysis system for the future, but this project was not funded. When Finland joined ESO in 2004 it was decided that, as part of its joining fee, it would provide a contribution in kind of software effort, amounting to more than 18 person-years over three years, devoted to starting to address this outstanding problem. The project was named Sampo (*www.eso.org/sampo*) after a mysterious artifact in the Finnish Kalevala legends. Although there is strong astronomical involvement with Sampo in Finland, the skills of the new team, who joined the project at the start of 2005 and are shown in the photo on the next page, are mostly in the areas of computer science. ESO is providing the project manager (RH) and project scientist (PM) who report to the ESO Data Management and Operations Division (DMD), which retains overall control of the project. In Finland the

team is coordinated by the National Coordinator (KV) and is based at two locations: CSC, the Finnish IT Center for Science, and the Observatory, University of Helsinki.

The scientific oversight comes from two bodies. The Finnish Astronomical Advisory Group (FAAG) represents the Finnish astronomical community and its interests and is chaired by one of us (JI). The ESO Faculty is represented by the Scientific Advisory Committee (SAC) that is chaired by the project scientist (PM).

In a parallel and earlier effort, an Opticon Network (3.6), with significant ESO involvement, is also looking at the question of future analysis systems and will also supply input to the project.



The current Sampo Team in Finland. From left to right: Sami Maisala, Kari Vasko, Johan Lindroos, Matti Anttila, Marko Ullgren and Tero Oittinen.

Sampo Project Goals

The project will run for three years (2005–2007). There are three main goals. Firstly the team must develop a clear idea of the science requirements for a data analysis system to satisfy the needs of the ESO community for the decade 2010–20. Among the questions to be addressed are: How should access to legacy code be provided? What should the glue needed to connect up existing software building blocks look like? What is the relative importance of the development of stand-alone Graphical User Interfaces, web applications or the provision of scripting languages with appropriate libraries and tools? At which level is interactivity needed? How does the data analysis environment provide services to the VO? Should tools for data organization be part of the system?

Secondly the project will identify risk areas and assess the technology required and test it on realistic astronomical data sets. Lastly the project will execute several pilot studies that illustrate critical steps along the road to a new system, validate the concepts and also produce significant tools of immediate value for the ESO community. This is a research and development project and can be regarded as a Phase A study.

The work will progress as a series of projects, each typically taking six months to a year, in which specific questions or problems are addressed. Initially, as the Finnish team was not familiar with astronomical software, a seminar was arranged in Helsinki and some small projects were selected to help them to learn about astronomical software. A more substantial project, PyMidas, is currently in progress and is described below.

The deliverables from these projects will be software tools that are pilot implementations addressing specific problems posed by the scientific oversight bodies listed above, complete with detailed documentation, or reports describing studies. As well as reporting to the ESO DMD, these projects will be written up and presented at the ADASS or other conferences and, where appropriate, presented as demonstrations. The model of annual cycles with demonstrations and meetings at the same times, over a period of several years, as adopted by the virtual observatory, will be followed, where appropriate.

When possible, projects will be chosen so that the resulting software tools are of direct value to the ESO community. As an example, Sampo might look at developing an architecture that would allow astronomers to use the UVES pipeline (that is based on ESO-MIDAS), newer pipelines based on CPL and other analysis tools in a convenient and uniform way. Such a development would be not only of benefit for the community but also be valuable for the operational teams who could then possibly use ESO-MIDAS pipelines not only on a local computer but also on a cluster. Another example might be addressing the questions posed by trying to connect up legacy software with data distributed through the virtual observatory in a convenient manner.

At the end of the three years a more detailed report will be presented to the ESO Head of DMD setting out conclusions from the projects and making detailed recommendations. Realistic resource estimates will be presented to allow the management to decide whether or not to proceed to a Phase B.

The First Pilot Project – PyMidas

As an initial project for the team we chose to develop an interface between the Python scripting language – which is now dominant in many areas of astronomy – and ESO-MIDAS (*www.eso.org/midas*), the main legacy data reduction and analysis system from ESO. This project is now well advanced and expected to be released in October 2005 and demonstrated at the ADASS 2005 (*www.adass.org*) conference. PyMidas allows any ESO-MIDAS command to be executed from Python using standard syntax, but also allows the developer to call on the full power of scripting in Python. ESO-MIDAS commands can be included in Python scripts – along with those from the many other Python-based systems such as PyRAF (*www.stsci.edu/resources/software_hardware/pyraf*) – and facilities from the ESO-MIDAS monitor (such as direct access to keywords, table elements and pixel values) will also be available directly from Python. This project will enhance the access to legacy code – and many valuable and trusted applications – whilst connecting the legacy with more modern technology as well as other systems.

Conclusions

A new software infrastructure will be needed to replace the ageing systems in current use if the ESO community is to be ready for the age of the virtual observatory and the new generation of instrumentation. The Sampo project, which has just started, aims to develop a detailed characterization of the requirements for future data reduction and analysis environments and to produce many useful components from pilot projects along the way. Some aspects of such a system are already apparent and we expect that a much clearer and more complete view will emerge over the three years of the project. An initial pilot project, PyMidas, is progressing well and will be made available in the autumn of 2005. More details about the Sampo project are available through the web pages at *www.eso.org/sampo* where there are links to other related documents. We very much welcome advice from the ESO community and encourage comments and questions to *sampo@eso.org*.

For many years the ESO community has asked for better tools for data analysis and data reduction. This goal has always been recognized by ESO as being very important, but a significant project could only be funded when Finland joined ESO. The Sampo project will be a major step forwards towards addressing these concerns.