

Requirements for a Future Astronomical Data Analysis Environment

Preben Grosbøl, Klaus Banse

European Southern Observatory, Germany

Doug Tody, William Cotton, Tim Cornwell

National Radio Astronomical Observatory, USA

Daniel Ponz

European Space Agency, Spain

Janne Ignatius

CSC, Finnish IT Center for Science, Finland

Peter Linde

Lund Observatory, Sweden

Thijs van der Hulst

Kapteyn Astronomical Institute, Netherlands

Vadim Burwitz

Max-Planck-Institut für extraterrestrische Physik, Germany

David Giarretta

Starlink Project, United Kingdom

Fabio Pasian, Bianca Garilli

Istituto Nazionale di Astrofisica, Italy

William Pence

National Aeronautics and Space Administration, USA

Dick Shaw

National Optical Astronomical Observatory, USA

Abstract. Most of the systems currently used to analyze astronomical data were designed and implemented more than a decade ago. Although they still are very useful for analysis, one often would like a better in-

terface to newer concepts like archives, Virtual Observatories and GRID. Further, incompatibilities between most of the current systems with respect to control language and semantics make it cumbersome to mix applications from different origins.

An OPTICON Network, funded by the Sixth Framework Programme of the European Commission, started this year to discuss high-level needs for an astronomical data analysis environment which could provide a flexible access to both legacy applications and new astronomical resources. The main objective of the Network is to establish widely accepted requirements and basic design recommendations for such an environment. The hope is that this effort will help other projects, which consider to implement such systems, in collaborating and achieving a common environment.

1. Why a New Common Software Environment?

The major issues facing astronomers who want to analyze their data may be summarized as follows:

- there exist many different systems but:
 - designed several decades ago,
 - largely incompatible with respect to scripting language used
 - complicated to share data between them (due to keyword semantics)
- current systems provide only rather limited interfaces to Web services, Virtual Observatories, archives and databases
- it is difficult to fully exploit available computer resources e.g. GRID

Due to old (often monolithic) designs, the current generation of systems will be difficult to upgrade to remedy the problems listed above. A new environment seems to be the better option. It would have to fulfill the following criteria:

- usage of important legacy applications from current systems
- easy and attractive for users to develop new tasks
- increase the ability to collaborate and of sharing of software
- support easy access to resources such as archives, Virtual Observatories, and GRID computing
- open, stable and well controlled interface specifications
- define minimum implementation but allow full featured version
- specify open, free base system but also provide interface to commercial software

2. High Level requirements

The success of a system/environment for data analysis is closely linked to the following attributes or points:

stability: Users have to be sure that their 'investment' in learning and using an environment also pays back in the long run.

well tested system: Although the system must be subject to its own regression tests, only actual users will be able to find subtle issues in application procedures.

availability of new state-of-the-art applications: This is only achievable if it's easy and attractive for users to develop new tasks in it.

support: Even with good documentation, high stability and excellent testing, the surrounding software world is not fixed demanding a constant although limited support.

collaboration: Modern science is often done in international teams. Features to make such collaborations as easy as possible must be provided.

up-to-date features: Error propagation and hypothesis testing are examples of features which a modern environment must support. Also easy access to compare results with models is important.

ease of installation: The environment must be trivial to install on most commonly used systems to ensure it can be used by 'normal' astronomers.

compatibility: Future data analysis will rely on access to Virtual Observatory data and Grid computer resources which must be fully supported.

3. What defines a new Environment?

The concept of what an environment is differs significantly depending on whom you talk to. In this context, an environment must have the following properties:

- execute tasks in a transparent way so that users can select easily the appropriate computing resources e.g. desktop, GRID or supercomputer
- provide a standard for parsing information between tasks
- offer a high level scripting language for flow control
- access data transparently no matter where they are located such as in databases or through Virtual Observatories
- make a set of standard services available such as display of data, define a standard interface to services provided

A UNIX system with a standard shell would satisfy many of these criteria although arguably not in an optimal way. The issue of defining a new environment may well be to find the minimum acceptable one.

4. From Dream to Reality

High level requirements, beautiful concepts and solid architectural designs are always good to have but not worth much if they are not realized in an actual implementation. The OPTICON Network is currently only funded to establish the high level definitions but several organizations and projects have concrete plans for implementations of such an environment for data analysis, for example:

ALMA: There is an explicit need for providing an off-line data analysis environment for ALMA data.

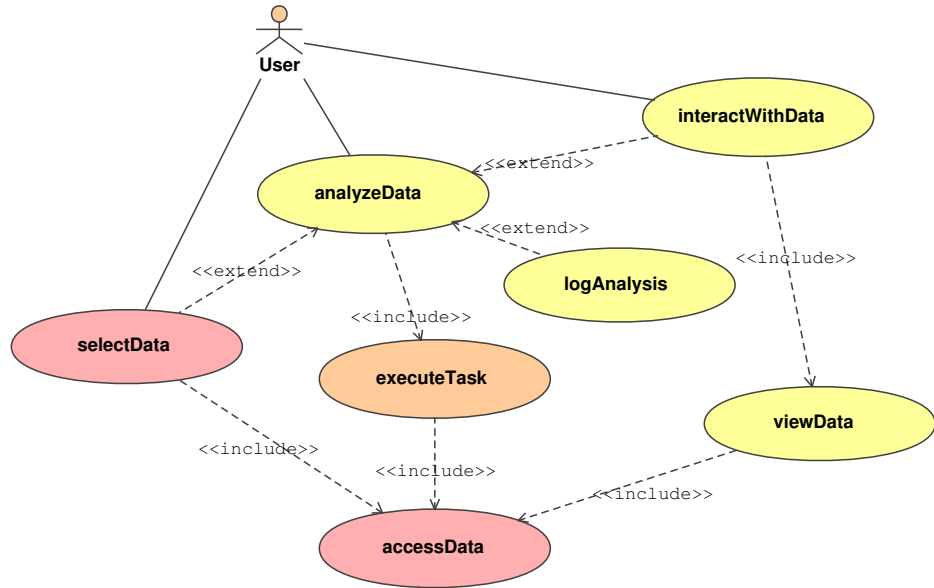


Figure 1. UML diagram of typical Use Case

ESO: The entry of Finland into ESO has made it possible to start a pilot project with the aim of providing better tools for the analysis of ESO data to its community.

The OPTICON Network offers a forum where general requirements and design concepts can be discussed as it includes people associate to these and other similar projects. There is a strong feeling within the Network that a global environment with identical interfaces must be established although implementation details may differ. However, in the end only the funding organizations can make this wish come through.

5. OPTICON Network on Future Astronomical Software Environments

OPTICON (Optical Infrared Coordination Network for Astronomy) includes a working group to discuss Future Astronomical Software Environments (FASE) for data analysis. The high level objectives for the FASE working group are:

- identify areas for which a common FASE is desirable and feasible
- establish high level requirements for a FASE
- draft interface and design recommendations for these areas
- learn from past experience

The network has currently 15 members from European institutes and 6 associated members from USA.

6. Top level User Scenario

Several user scenarios were considered including data analysis and development of tasks by both teams and individuals. A typical case for an astronomical user is shown as an UML diagram in Fig. 1. The Use Cases are summarized as:

analyzeData: provides a high level scripting language for controlling the flow of the data analysis that is execution of specific tasks

selectData: lets the user view possible data sets and select appropriate ones

interactWithData: offers the ability to interact with the data by graphics or other means

logAnalysis: logs all actions and results providing the user with a comprehensive view over what has been done

executeTask: executes individual tasks either locally or in a Grid context

viewData: displays data to the user

accessData: provides transparent access to data either locally or at remote sites through the Virtual Observatory interface.

Several of these Use Cases (e.g. `accessData`, `selectData`) will be based on the concepts and tools provided by Virtual Observatories while others (e.g. `executeTask`) will use parts of Grid technology.

7. How can you give your input?

It is important that potential users of the environment participate as much as possible in laying down its requirements, both scientific and technical. Although one would not be able to satisfy all wishes, it is essential to have a detailed view of what the community would like to see and work with. The two main channels for this are discussions on:

- the TWiki: <<http://archive.eso.org/opticon/twiki/bin/view/Main>>
- the Majordomo e-mail list <fase@eso.org>

Acknowledgments. This work is supported by OPTICON which is funded by the European Commission under Contract no. RII3-CT-2004-001566.