eSTAR: Telescopes and Databases as a Single Information Grid

Tim Naylor[1], Iain Steele[2], Dave Carter[2], Alasdair Allan[1], Jason Etherton[2], and Chris Mottram[2]

[1] School of Physics, University of Exeter, Stocker Road, Exeter EX4 4QL, U.K
[2] Astrophysics Research Group, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD, U.K

Abstract. We present a sketch of a system which consists of a Grid of databases and telescopes within which intelligent agents assess the information available to them, and as a result request the telescopes take new data. These systems allow programmes to be carried out much more efficiently than is currently possible, and their speed of reaction enables novel programmes that are unavailable to current technology. We then present eSTAR, a currently functioning technology prover for such systems which use computational grids software (Globus) to create an information Grid.

1 Introduction

Not so long in the future, in a galaxy not far, far away astronomers had built a Virtual Observatory. But a small, determined rebel group continued to make new observations! "What has the mighty federation of databases ever done for us?" they cried.

I began my talk at the AVO conference with this spoof of the Star Wars opening sequence (and with a minor dig from Monty Python's Life of Brian) to make a very serious point about the aims of the Virtual Observatory initiatives. Whilst the end result may be a seamless federation of databases, there will be a significant number of astronomers (possibly the majority), still involved in data acquisition. That data acquisition process often requires the use of databases, and increased use of databases will clearly benefit the process. What I hope to persuade you in this contribution is that making a seamless interface between telescopes and databases, indeed making telescopes look like databases will bring enormous benefits. But, there is functionality telescopes need that databases don't, so a degree of planning is needed at this stage to ensure convergence in the paradigms used for grid enabled telescopes and databases.

What can be achieved is best seen by considering a couple of possible science drivers. So, imagine we have built a system which not only has unified access to databases and telescopes, but also to the astronomical literature. In addition the system has intelligent agents which can interpret the results. One can view this system as a unified information grid, within which these intelligent agents live.

2 Driver 1: The Space Density of Dwarf Novae.

Dwarf novae are semi-detached interacting binaries where a Roche-lobe filling low-mass star loses mass to a white dwarf. There is a thermally unstable accretion disc around the white dwarf. When disc is in outburst it is almost a standard candle, which outshines all the other sources of light in the system. Hence acquiring a sample which is magnitude limited when in outburst is almost identical to obtaining the volume limited sample required for measuring the space density. The problem is that dwarf novae spend only $\sim$10% of their time in outburst, the remaining time they are $2-5$ astronomical magnitudes fainter. Thus conventional, one epoch sky surveys are not a good way to tackle this problem.

Instead we could imagine a programme where an intelligent agent (IA) compares every CCD field taken in the world with the SuperCosmos scanning of Schmidt plates. The IA could then carry out the following tasks.
1) Objects which brighten above some fixed magnitude (say 16th) are cross-correlated with the SIMBAD stellar database.
2) Those which SIMBAD shows are known dwarf novae are noted, and those which are known not to be dwarf novae discarded.
3) For the previously unknown objects, archive data are searched to identify their nature from their lightcurves.
4) Where the archive data are too sparse, the IA requests new observations, building up a lightcurve, ensuring that new observations are placed at critical times to define whether or not the object is a dwarf nova.
5) The IA catalogues objects it is certain are dwarf novae, and requests a parallax measurement from HST to confirm the distance.

After this entire process is complete, the astronomer returns from a (presumably quite extended) lunch break, and is ready to write a paper.

3 Driver 2: What was that?

The above example is a science problem we could already solve, but the system proposed here could do it faster and cheaper than more traditional methods. However, the proposed system offers a speed of action which not currently possible, opening new possibilities. Consider the following example, with timings which I will justify later.

02:11:03UT. The shutter closes on a wide-angle CCD image of Centaurus taken using a large camera lens.

02:12:30UT. The data have been processed, and a list of magnitudes and positions is made available on the astronomical data grid.

02:12:45UT. An astronomer's intelligent agent has compared the source list with archive data, finding a bright object.

02:13:00UT. In response to the IA's request for confirmation a small telescope slews to acquire another image. Whilst waiting for this the IA queries SIMBAD and discovers there is no known variable at this position.

02:15:06UT. The new image confirms the object, so the IA requests spectroscopy from the Liverpool Telescope. During this waiting period the IA pulls all the other available data and papers.

02:22:34UT. The IA recognises that the spectrum is odd. There hasn't been a γ-ray burst, but VISTA shows a very faint red object mentioned in a paper last year.

02:22:50UT. Less than 11 minutes after the initial exposure ended, the IA has run to the end of its pre-programed "understanding", so it wakes up an astronomer.

That astronomer already has a unique dataset, and a great deal of the preliminary work of searching databases and literature has been done. But now some creative thinking is required. The majority of the time delay has been introduced by exposure time; I have assumed that the data pipelines can reduce the data quickly, and will discuss below how rapidly the IAs can react.

4 The eSTAR Project.

So how close are we to achieving a system like this? The eSTAR project (e-Science Telescopes for Astronomical Research) is being funded by the UK's Department of Trade and Industry, initially as an e-Science demonstrator project. We use Meade LX200 and ETX telescopes with Santa Barbara Instruments Group (SBIG) or Apogee CCD cameras. These are chosen mainly because they have a simple programming interface, allowing us time to concentrate on the novel issues of the problem. At the moment the system works across the network as a daytime demonstration, slewing the telescopes and taking CCD images, but then sending fake data to the IAs. We plan complete tests taking real data later in 2002.

Our main design criterion was to have something that works in an analogous way to a computational grid, with no overall supervisor. Each telescope has its own scheduler, which can talk to the IAs using the Robotic Telescope Markup Language (RTML – see F.V. Hessman's paper in these proceedings) and transported over Globus. The IAs also have to communicate with databases such as SIMBAD and the Astrophysical Data Service (ADS), and datasets such as the US Naval Observatory's A2 astrometric catalogue (USNO-A2), and the Digital Sky Survey (DSS). This is done using WWW interfaces. Such a system should be eminently scalable, with many telescopes (we have already tried four) and many intelligent agents. The intelligent agent we see as carrying out a specific scientific programme, and probably resides on the astronomer's desktop computer.

A typical sequence begins with the IA carrying out a Globus resource discovery (LDAP), finding each telescope and its associated instrument suite. Next it asks which telescopes could carry out a particular observation, each node returning a score, which encodes the predicted quality the observation. The IA then selects a telescope it wishes to carry out the observation, and then requests it, with the scheduler placing it in the observing queue. Once the observation is complete the raw and reduced data are made available to the IA. Our current

IA can compare the new observations with USNO-A2, finding variable stars, and carrying out a search at these co-ordinates in SIMBAD. The IA will then grab a DSS image of the area. If the variable star turns out to be a dwarf nova, then the IA will request new observations, in addition to acquiring a complete set of references for the object. Having created this system, we were pleasantly surprised to discover how fast the cross-correlation runs, normally about six seconds, including recovering data from three web sites. This shows that the IAs would not be the limiting speed factor in a rapid reaction system.

If a system similar to this is to become widespread, the idea of requesting observations is important. It allows telescopes to join the system with outside intelligent agents being, perhaps, only a small part of the observations the scheduler has to deal with. Furthermore, telescopes can reject requests from an IA if, for example, the astronomer has not got time allocated on the telescope. In the longer term we hope to use the authentication and authorisation of users of grid resources, as well as accounting for the resources used. Observatories will want to be sure that only authorised people are using their resources. Equally important, though, is that such a system will be easier to create if databases and telescopes have a common interface. Databases are only special cases of telescopes, that can return data instantly, rather than after a time delay.