AstroGrid: the UK's Virtual Observatory Initiative

Robert G. Mann

Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, United Kingdom

for the AstroGrid Consortium[1]

Abstract.

AstroGrid is the UK's Virtual Observatory (VO) initiative. It brings together the principal astronomical data centres in the UK, and has been funded to the tune of ~£5M over the next three years, via PPARC, as part of the UK e–science programme. Its twin goals are the provision of the infrastructure and tools for the federation and exploitation of large astronomical (X-ray to radio), solar and space plasma physics datasets, and the delivery of federations of current datasets for its user communities to exploit using those tools.

Whilst AstroGrid's work will be centred on existing and future (e.g. VISTA) UK datasets, it will seek solutions to generic VO problems and will contribute to the developing international virtual observatory framework: AstroGrid is a member of the EU-funded Astrophysical Virtual Observatory (AVO, see **www.eso.org/projects/avo** and Quinn 2002) project, has close links to a second EU Grid initiative, the European Grid of Solar Observations (EGSO, see **www.mssl.ucl.ac.uk/grid/egso/**), and will seek an active role in the development of the common standards on which the international virtual observatory will rely.

In this paper we shall primarily describe the concrete plans for AstroGrid's one-year Phase A study, which will centre on: (i) the definition of detailed science requirements through community consultation; (ii) the undertaking of a "functionality market survey" to test the utility of existing technologies for the VO; and (iii) a pilot programme of database federations, each addressing different aspects of the general database federation problem. Further information on AstroGrid can be found at **www.astrogrid.ac.uk**.

---

[1]Andy Lawrence, Clive Davenhall, Bob Mann (Institute for Astronomy, University of Edinburgh), Richard McMahon, Mike Irwin, Nic Walton, Guy Rixon (Institute of Astronomy, University of Cambridge), Mike Watson, Julian Osborne, Clive Page (Department of Physics & Astronomy, University of Leicester), Peter Allan, David Giaretta, Chris Perry, Dave Pike, John Sherman (Space Data Division, Rutherford Appleton Laboratory), Fionn Murtagh (School of Computer Science, Queens University Belfast), Louise Harra, Bob Bentley, Keith Mason (Mullard Space Science Laboratory, University College London), Simon Garrington (Jodrell Bank Observatory, University of Manchester).

1. Introduction: The Origins of AstroGrid

In common with other virtual observatory initiatives, AstroGrid arose from a fortuitous convergence of the needs of astronomy, the capabilities of computing hardware, the research interests of computer scientists, and the willingness of funding agencies to support them all. In 2000, PPARC's Astronomy Long-Term Science Review[1] panel placed *Computing and Databases* on its "Priority 1" list, in recognition of the central role that the coming generation of survey archives will play in UK astronomy in the next 10-15 years. The funding to develop such resources became available later that year, when the UK Office of Science & Technology (OST) announced a new £98M, three–year programme for "e–science", which it defined as *"the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet"*. In addition to allocations to individual Research Councils, funding was provided for an e–Science Core Programme[2] to *"develop and broker generic technology solutions and generic middleware to enable e–Science and form the basis for new commercial e–business software"*. The AstroGrid consortium formed in the run–up to the OST announcement, as it became clear that the major data centres in UK astronomy, solar and space plasma physics shared similar database volume and federation problems, in the face of major forthcoming projects, and a similar keenness to seek solutions to them based on the Grid computing concept.

2. The Goals of AstroGrid

AstroGrid will focus on short term deliverables, both relevant application tools and the federation, by the data centres that manage them, of key UK–held astronomical, solar and space plasma datasets. This work will be split into a one–year Phase A experimental study and a two–year Phase B implementation: we discuss the former in Section 3, but the details of the latter are as yet incomplete, as they rely on the results of our Phase A studies and on the coordination that develops between the various VO/Grid initiatives in the next year or so.

3. AstroGrid Phase A: Experimental Study

AstroGrid's initial activities fall under three broad headings, as follows:

3.1. Developing Science Requirements

The broad science case for the VO has been well aired (for example in Brunner 2002, and in the conference proceedings of Brunner, Djorgovski & Szalay 2001, and Banday, Zaroubi & Bartelmann 2001), but AstroGrid needs detailed science requirements, tailored to the specific priorities of the UK astronomical community. These will be derived via community consultation, with open meetings and visits to UK astronomy groups, followed up by brainstorming sessions where key

---

[1] `http://www.pparc.ac.uk/Rs/Pp/Sp/AstronomyLTSR.asp`

[2] `http://www.research-councils.ac.uk/escience/`

project staff and invited participants will distill inputs from the community into a prioritised set of requirements which will delineate the scope of the AstroGrid programme.

3.2. Testing Technology

In tandem with the derivation of its science requirements, AstroGrid will undertake a "functionality market survey", assessing how existing techologies meet the needs of the virtual observatory. This programme is divided into three parts, matching the workpackages of the AVO's Technology work area (Quinn 2002).

Data–Grid Technology: Evaluating Grid components and related technologies, such as the Globus Toolkit, XML, Storage Resource Broker (SRB), JINI, CORBA, SOAP, etc, via test deployments of trial data–grids (*e.g.* Rixon et al. 2002).

Compute/Store Technology: Developing and performing benchmarks tests that realistically simulate AstroGrid operations, to assess the availability, performance, robustness, scalability and cost of different hardware systems, and to produce recommendations as to what should be procured for AstroGrid's Phase B implementation.

Database Technology: Establishing the functional requirements of the database technology that will under–pin the VO, and assessing how a range of existing database management systems meet those needs.

Some of this activity will be coordinated with other e–science projects, where it covers basic Grid issues (user authentication and authorization, resource discovery and allocation, data transfer, etc) that are common to all, but other topics, such as the crucial issue of metadata, will be astronomy-specific.

3.3. Undertaking a Pilot Programme of Database Federations

This third main strand of Phase A has two motivations. Firstly, by providing hands–on experience of real problems, it will help in the assessment of software and hardware capabilities and the identification of functional requirements, and, secondly, by delivering early prototype tools to our user communities, it will help them see what the VO can make possible, thereby facilitating a more informed definition of science requirements for AstroGrid.

With these motivations in mind, we have selected a set of five pilots, which have been chosen both because their differing data types mean that they address different aspects of the general database federation problem, and because they will deliver VO prototypes to different user communities:
(i) *Large object catalogues – optical/NIR*: addressing scalability issues by federating SuperCOSMOS and Sloan catalogues in the SDSS Early Data Release (Stoughton et al. 2002) region, and catalogues from the optical and near-IR survey instruments on the INT (WFC and CIRSI, respectively);
(ii) *Event lists – X-ray*: generating new data products by combining selected Chandra and XMM–Newton datasets, thereby assessing the completeness of the OGIP standard as a metadata model for the VO;

(iii) *Fourier data – radio*: integrating existing radio images via registration and overlay, and investigating the generation of images on–the–fly from stored visibility data, according to user–specified routines;
(iv) *Image and movie libraries – solar*: developing visualisation tools and procedures to ease the selection of datasets for detailed analysis using quick–look data products in the form of synoptic images and movies;
(v) *Time series data – space plasma*: assessing the metadata and procedures required for the integration of time series data with different sampling profiles taken by different instruments.

4. AstroGrid Phase B: Implementation

The results from the Phase A studies will be combined with the picture of coordination between VO and Grid initiatives internationally, to define the detailed plan for AstroGrid's Phase B implementation of a working astronomical data–grid, integrated into the developing global VO and capable of accommodating the coming generation of major UK–held datasets.

5. Summary and Conclusions

A sophisticated, global Virtual Observatory will take longer to develop than the 3–5 year lifetimes of the currently–funded VO projects. Some parts of this process are necessarily *top-down* work, such as reaching agreement on astronomical metadata standards and Grid protocols, but there must be a parallel track of *bottom-up* R&D, if the VO's user communities are to become engaged, its funding agencies kept interested, and its potential start to become realised through the performance of new, better, and more efficient science.

AstroGrid will contribute fully to the coordinated development of standards, but it allies itself closely with the pragmatic approach of starting to get the VO working through hands–on experimentation. The e–science programme is expected to continue into the next UK Govt spending round, so the ∼£5M awarded to AstroGrid for the next three years – together with its role within the AVO (Quinn 2002), and middleware development to the benefit of astronomy undertaken by the e–Science Core Programme – should be regarded only as the UK's initial contribution to developing the global Virtual Observatory.

References

Banday, A.J., Zaroubi, S., & Bartelmann, M., 2001, (eds.) Mining the Sky (Berlin: Springer)

Brunner, R.J. 2002, this volume, [O-1]

Brunner, R.J., Djorgovski, S.G., & Szalay, A.S., 2001, (eds.) Virtual Observatories of the Future (San Francisco: ASP)

Quinn, P.J. 2002, this volume, [O-19]

Rixon, G.T., et al., 2002, this volume, [D-7]

Stoughton, C., et al., 2002, AJ, submitted.