AstroGrid-2

A proposal to PPARC from the following lead investigators:

Institute	***Lead Investigator***
School of Computer Science, Queens University of Belfast	Prof Fionn Murtagh
Department of Physics, University of Bristol	Dr Malcolm Bremer
Institute of Astronomy, University of Cambridge	Dr Richard McMahon
Institute of Astronomy, University of Cambridge	Dr Nicholas Walton
Institute for Astronomy, University of Edinburgh	Prof Andrew Lawrence
School of Physics, University of Exeter	Prof Tim Naylor
School of Computing, University of Leeds	Prof Ken Brodlie
Dept of Physics and Astronomy, University of Leicester	Dr Mike Watson
Dept of Physics and Astronomy, University of Leicester	Tony Linde
Mullard Space Science Laboratory, University College London	Dr Louise Harra
Jodrell Bank Observatory, University of Manchester	Dr Simon Garrington
Institute of Cosmology & Gravitation, University of Portsmouth	Dr Robert Nichol
Space Science and Technology Department, Rutherford Appleton Laboratory, CCLRC	Dr Peter Allan

Table of Contents

Table of Contents

0. EXECUTIVE SUMMARY

We propose a programme of work to develop a future Virtual Observatory infrastructure for the UK that delivers powerful analysis facilities, is matched to key facilities and missions, is integrated into the European scene, and backs UK data centres in international competition.

The Virtual Observatory is conceived as a set of standards and a software framework that allows creative diversity in publishing data services and writing software tools. The current AstroGrid project is on track to produce a first working version of this ideal. In addition, we will populate the system with selected current UK databases, and provide a limited set of user tools. However, there is a clear need for a follow−on project − to extend and improve the infrastructure, to take advantage of further waves of technology, and to establish UK Data Centres as competitors in the new world, especially as we head towards competition over servicing key missions such as ALMA, GAIA, or JWST.

We have designed a broad but structured project to address these new priorities. We aim to :

- Expand, revise and improve the core infrastructure
- Provide a suite of science user tools, and set up a system for responding to personalised user needs
- Research and deploy new automated resource discovery techniques − AstroOntology and Intelligent Agents
- Research and deploy techniques in Grid technology, visualisation and datamining, creating an AstroGrid Data Exploration Framework (ADEF)
- Create a UK Data Centre Alliance, and provide co−ordinated effort to take up VO and Grid technology
- Develop dedicated outreach software to allow schools and the public to access the Virtual Observatory

These goals will be pursued within the context of AstroGrid's role in the EURO−VO programme that has been proposed to FP6. Specifically, AstroGrid is expected to lead a distributed collaboration known as the "VO Technology Centre (VOTC)", as well as its constituent data centres being members of a Europe−wide Data Centre Alliance (DCA). Our work programme will also be co−ordinated with the general UK e−science programme, and the international VO programme, through the IVOA.

The AstroGrid−2 project will be delivered by an expanded version of the AstroGrid consortium, involving eleven institutions, over fifty personnel, and a broad range of astronomical and computer science expertise. Ten key science and management staff are named in the proposal, and a further thirty developers and RAs are budgeted, with precise deployment to be determined later. The total cost from April 2004 to December 2007 is £9.2M.

1. INTRODUCTION

1.1 AstroGrid and the Virtual Observatory (VO) concept

The aim of AstroGrid and other VO projects is to make doing astronomy faster, more effective, and more economic, by standardising the data analysis process and by freeing the astronomer from many mundane tasks. It also has the potential to influence the discovery process in astronomy in a dramatic way − by encouraging new styles of data−intensive exploratory science, by removing interdisciplinary barriers, and by encouraging the pooling of resource and the formation of distributed collaborative teams. We also expect that it will be a liberating force in that the resource available to astronomers will become almost independent of their location.

The VO is conceived as a system that allows users to interrogate multiple data centres in a seamless and transparent way, which provides new powerful analysis and visualisation tools within that system, and which gives data centres a standard framework for publishing and delivering services using their data. This is made possible by standardisation of data and metadata, by standardisation of data exchange methods, by the use of Registries which list available services and what can be done with them, and by the creation of *workflow tools*. The vision is not one of a fixed specific software package, but rather one of a *framework* which enables ***data centres*** to provide competing and co−operating *data services*, and which enables ***software providers*** to offer a variety of compatible *analysis and visualisation tools* and *user interfaces*.

The standard VO framework is emerging on an international scale, but AstroGrid will produce the first working set of infrastructural software components, including a user interface and basic tools. AstroGrid is also however a ***consortium of data centres and software providers***, and so expects to *publish* data services and analysis tools to the VO framework, to make a complete working implementation. The consortium will pool resources, including key UK databases, storage, and compute facilities, and will operate them co−operatively using Grid middleware (OGSA).

1.2 The case for AstroGrid−2.

The AstroGrid project completes in December 2004 but the Virtual Observatory will not be finished. Our aim is to have a working set of infrastructural software components, but the analysis tools will be very limited, and the background technology and international standards will continue to evolve. Furthermore our focus on a framework to which data services are published also places a new burden and importance on professional data centres in astronomy, requiring effort to deploy VO and Grid technology and generally to raise their game, to be ready to compete for position in key upcoming missions and facilities such as ALMA, GAIA, JWST etc. Each of those pipeline/archive projects will bid for their own resources, but we feel there is a strong case to put the data centres concerned in a competitive position. Finally, provision for advanced tools and techniques, eg datamining, visualisation, and automated resource discovery, has only just begun, and requires substantial R&D as well as collaboration with computer scientists specialising in these areas.

Our idea then is to complete the existing AstroGrid project, but then to begin a *new* project with three main themes. (i) Refreshing and improving the VO infrastucture, especially user tools; (ii) Research and development in advanced new techniques; and (iii) VO uptake and backbone support for a Data Centre Alliance.

1.3 International Context : IVOA and Euro−VO.

There are currently 12 VO projects worldwide, linked through an International Virtual Observatory Alliance (IVOA) which acts both as discussion forum and as a semi−formal standards development and approval body. The IVOA is dominated by US−NVO, the European AVO project, and AstroGrid, and we expect this to continue. US−NVO is funded to 2006, but AVO, like AstroGrid, completes in 2004. AstroGrid is a partner in a new venture known as "Euro−VO" which we are seeking to create through a series of FP6 proposals. We have carefully aligned Euro−VO and AstroGrid−2 goals − VO core extension, science user tools, data mining, visualisation, and data centre support. The Euro−VO project aims to create three organisations − a VO Facility Centre (VOFC), a Data Centre Alliance (DCA), and a VO Technology Centre (VOTC). ESO is hosting the VOFC, and CDS Strasbourg is leading the DCA network. The VOTC is intended as a distributed collaboration, but with acknowledged AstroGrid leadership. AstroGrid's aim is to be the technology powerhouse of Europe. This will be credible if we can match Euro−funding with UK funding. In particular Euro−VO has deliberately created "co−ordinator reserves" which will only be allocated to partners after a year's study phase.

1.4 Long Term Implications

In a direct sense, AstroGrid−2 does not have a "whole−life−cost" which we should present here − like the original AstroGrid, it is designed to be self−standing and finite. Any follow−on work will be separately proposed and justified at the appropriate time. Furthermore, part of the VO philosophy is that the long−lived things are the standards and protocols, so that any infrastructure implementation can be thrown away and a new one written. Nonetheless, there are several ways in which we think PPARC should make planning provision for long term support in this area :

(a) Data centres, and expert service centres more generally, will play a crucial role in astronomy for many years to come, but this role is becoming increasingly difficult and professionalised. A small number of UK data centres should be established as "well found labs" so as to be competitive on the international scene.

(b) AstroGrid software needs to be maintained until it is superseded. We are designing to minimise this, for example by making as much as possible browser−mediated to avoid "platform madness", but there is still an irreducible minimum.

(c) The international VO scene is clearly going to be with us for a long time, and some persisting body may be needed to participate in continuing standards development and other IVOA matters. However, much of this may be routed through the Euro−funded VOFC.

1.5 The current proposal.

To achieve our new aims, we have developed an expanded consortium, adding four new partners − Leeds, Bristol, Exeter, and Portsmouth. This represents expertise in intelligent agents (Naylor), visualisation (Brodlie, Bremer), and datamining algorithms (Nichol). In addition, we will work with the eDIKT team at NeSC/EPCC, who will provide a developer from their funds to match one of ours in collaborative datamining work. We have also identified a set of other "associated projects" which will be independent of AstroGrid−2, but with which we will work closely. As well as Euro−VO and the UK e−science programme, with which we intend to be fully integrated, this will include SSVO (solar system) and TVO/VirtU (theory) as well as key pipeline/archive projects (VEGA and ALMA).

AstroGrid includes solar and STP requirements in its work, and most of what AstroGrid is building is completely generic. In this new round, it is intended that AstroGrid−2 will continue to provide *generic* infrastructure for solar work, but that SSVO will develop its application specific software. (We expect however to collaborate closely on visualisation problems). SSVO will also bid for its own data centre support. This AstroGrid−2 proposal therefore bids only for optical/IR/X−ray/radio centres at Cambridge, Edinburgh, Jodrell Bank, and Leicester.

The rest of this proposal is divided as follows. (i) Science Requirements capture for AstroGrid−2. (ii) Infrastructure extensions, including re−factoring, new components, and development of basic user tools. (iii) Our involvement in the international standards programme. (iv) Research and development in new technologies − astro−ontology, visualisation, intelligent agents, the idea of an AstroGrid Data Exploration Framework, and continuing Grid research. (v) The case for support for the UK Data Centre Alliance. (vi) Software development for VO outreach. (vii) A description of associated projects. (viii) A summary project plan, with management plan, budget, risk analysis etc. We also provide appendices with some more detailed information including relevant publications.

2. ASTROGRID

2.1 Introduction

AstroGrid is one of three major world−wide projects (along with European AVO and US−VO projects) which aim to create an astronomical ***Virtual Observatory***. The Virtual Observatory will be a set of co−operating and interoperable software systems that:

- allow users to interrogate multiple data centres in a seamless and transparent way;
- provide powerful new analysis and visualisation tools;
- give data centres a standard framework for publishing and delivering services using their data.

The long term vision of the Virtual Observatory is not one of a single software package, but rather of a framework which enables data centres to provide competing and co−operating data services, and software providers to offer compatible analysis and visualisation tools. The first priority of AstroGrid, along with the other VO projects worldwide, is to develop this standardised framework to allow such creative diversity.

AstroGrid officially began in September 2001 and was conceived of as a two−phase project. Phase A, which completed in December 2002, included research into emerging web and grid technologies and the building of prototypes which tested the feasibility of desired aspects of the VO. Phase A was fully documented in a Report to the AstroGrid Oversight Committee (AGOC) and the Grid Steering Committee (GSC) (see below).

Phase B began in January 2003. Its goal is to develop, within the period to December 2004, the components outlined in the Phase A Report so as to realise a working VO infrastructure giving access to prime UK data sources. The development process we are following has the goal of releasing working products every quarter (see 2.4 below).

Each iteration within Phase B will incrementally add features to existing components and create new components within the infrastructure. The features to be added each quarter are determined by a panel made up of scientists, technologists and institute representatives to ensure that each iteration not only follows the overall development plan but also addresses issues which have been raised by those communities and through software testing. The whole process is overseen by the AstroGrid Oversight Committee which reviews progress every six months.

2.2 Aims and Goals

In summary, the ***scientific aims*** of AstroGrid are:

- to improve the quality, efficiency, ease, speed, and cost−effectiveness of on−line astronomical research
- to make comparison and integration of data from diverse sources seamless and transparent
- to remove data analysis barriers to interdisciplinary research
- to make science involving manipulation of large datasets as easy and as powerful as possible.

And our top−level ***practical goals***:

- to develop, with our IVOA partners, internationally agreed standards for data, metadata, data exchange and provenance
- to develop a software infrastructure for data resources
- to demonstrate a physical grid of resources across key data centres
- to develop a searchable, self−replicating and easily manageable resource registry
- to implement a working Virtual Observatory system based around key UK databases and of real scientific use to astronomers
- to provide a user interface to that VO system
- to provide, either by construction or by adaptation, a set of science user tools to work with that VO system
- to establish a leading position for the UK in VO work

2.3 AstroGrid Collaborators and Personnel

AstroGrid is a collaboration between seven institutes, each of whom provides one of the ***Lead Investigators***:

- Prof Fionn Murtagh, School of Computer Science, Queens University of Belfast
- Dr Malcolm Bremer, Department of Physics, University of Bristol
- Dr Richard McMahon, Institute of Astronomy, University of Cambridge
- Prof Andrew Lawrence, Institute for Astronomy, University of Edinburgh
- Dr Mike Watson, Dept of Physics and Astronomy, University of Leicester
- Dr Louise Harra , Mullard Space Science Laboratory, University College London
- Dr Simon Garrington, Jodrell Bank Observatory, University of Manchester
- Dr Peter Allan , Space Science and Technology Department, Rutherford Appleton Laboratory, CCLRC

Day to day management of the project is in the hands of:

- ***Project Lead***: Prof Andrew Lawrence, Institute for Astronomy, University of Edinburgh
- ***Project Scientist***: Dr Nicholas Walton, Institute of Astronomy, University of Cambridge
- ***Project Manager***: Tony Linde, Dept of Physics and Astronomy, University of Leicester

In addition there are, as of June 2003, 23 other people employed (mainly full−, some part−) time on the project as researchers or developers.

2.4 Progress To Date

Key deliverables and milestones within Phase A include:

- defined an extensive set of science problems which a VO could make easier to solve
- selected ten *key* science problems as AstroGrid drivers
- created detailed use cases which identified the basic components of a VO infrastructure
- carried out extensive pilot studies demonstrating the feasibility of certain technology approaches
- established a working grid across several institutions (using Globus V2 toolkit)
- selected as OGSA−DAI 'early adopter'
- developed the web service−based ACE (astronomical catalog extractor)
- deployed ACE within the AVO first−light demo before press and scientists

AstroGrid is the clear technology leader within the group of European VO projects and, following on from our deployment of a set of ground−breaking project collaboration tools are world leaders in scientific project communications (many new projects, especially in the VO world are adopting one or more of our tools).

Phase B began in January 2003 with the first iteration completed and component software (prototype portal and registry) delivered as scheduled on 31−March. Iteration 02 is on schedule to complete on 30−June. This suite of components will demonstrate a complete, though functionally very limited, virtual observatory comprising:

- a web−based portal
- simple workflow construction
- registry query to identify and find relevant datasets
- submission of a standardised query to multiple datasets
- notification that a job has completed
- return of results to a private server−based storage area
- viewing of those results from the portal

This software will be demonstrated as part of the IVOA stand at the General Assembly of the IAU, which will be held from 13th to 26th July, 2003 in Sydney, Australia.

In parallel to our VO development efforts, we are extensively researching Grid technologies in order to integrate them into our VO components. As part of this effort, we are developing, in conjunction with the Australia−VO project, a grid−based set of tools which will extract cubes of data (eg from velocity maps from radio telescopes or imaging spectrographs or galaxy distributions from redshift surveys) from large scale databases, ship them to a beowulf cluster where they will be volume rendered and then ship the images in real−time to the user.

2.5 Project Documentation

The science goals of AstroGrid are explained in two documents:

- The 'AstroGrid Ten': http://wiki.astrogrid.org/bin/view/Astrogrid/ScienceProblems
  lists the science problems which the AstroGrid VO infrastructure will make easier to solve
- Science Milestones: http://wiki.astrogrid.org/bin/view/Astrogrid/ScienceMilestonesR01
  the ability to tackle aspects of the above problems will provide a measure of project progress

More detailed information on progress to date can be viewed in the following reports and documents:

- Phase A Report: http://wiki.astrogrid.org/bin/view/Astrogrid/PhaseAReport
- Grid Steering Committee Presentation:
  http://wiki.astrogrid.org/pub/Astrogrid/GridSteeringCommittee/GSC−Feb03−PLtalk.ppt
- Oversight Page: http://wiki.astrogrid.org/bin/view/Astrogrid/OversightPage
  This page is always up to date with the latest progress information
- Phase B Structure: http://wiki.astrogrid.org/bin/view/Astrogrid/PhaseBStructure
  This page has links to individual component pages from which designs, plans and other documents can be viewed

2.6 Technical Summary

Our vision of a component−based VO where interoperability is assured through internationally agreed standards is rapidly coming to pass. Our involvement in the International VO Alliance (IVOA) is driving standards development, particularly in areas critical to AstroGrid (VO Query Language and Registry). Other IVOA projects are building prototypes and demonstrations which are standards−compliant, even though the standards are still undergoing rapid evolution, and AstroGrid is testing those standards within a working (though still developing) VO infrastructure. This mode of operation is likely to continue for the rest of the AG1 project.

This structure is still evolving as the standards efforts continue and as we learn from testing the infrastructure as it is developed but the major components will definitely exist in the final VO. These are:

Portal

The portal will be a server−based web−delivered component through which all VO user interfaces are delivered. Each component which requires user interaction will do so through a separate page. These pages will conform to certain xml−based protocols so that software can interact with the user through the portal. Thus, any developer can add their own product's functionality into a portal simply by exposing it to this protocol.

Community

This will provide the ability to construct an online community consisting of both individuals and groups. A resource centre can then assign permission to use its resources (data sets, services etc.) to one or more groups within a community instead of having to name the individuals themselves. Within a community, the administrator of the community can assign rights to individuals and groups, including the right to add members and create groups.

AstroPass

The AstroPass is our name for a central server which will store user credentials, in much the same way as the Microsoft Passport scheme. Initially, we will simply allow a user to identify themself with a username/password combination but later will accept the upload of user certificates. A user can determine how much or how little information is exchanged with other VO portals.

Workflow

This component will provide the core functionality of the VO, enabling the user to construct complex workflows, adding queries and data analyses, uploading and downloading data and algorithms, rendering the results as tables or images etc. The workflow engine will be intelligent enough to determine when the output of one component does not match the input of another and to suggest a translator. Functional flows may be forked and joined under specific conditions. At points, the user can insert breakpoints where the flow will be interrupted to allow the user to undertake inspection of the intermediate data or carry out some manual process. As well as constructing complex workflows, the user may also simply enter single queries to registries or datasets.

Job Control

This component allows the user to inspect the status of a job which has been submitted and, if desired, change the run parameters of that job. Initially we will develop a simple job monitoring portlet but will add the ability to interrupt and change jobs later.

Registry

In AstroGrid, the registry will be fine−grained: so will contain detailed information about resources, allowing the user to locate a specific resource without queries having to be sent to the resource centres themselves. In the first stage, we will store simple dataset information (wavelength, number of entries, data table schema) and allow simple queries. This will expand as we add more and more data and types of data and features to manage and query the registry.

MySpace

MySpace will provide users and components with a virtual space in which data can be temporarily and permanently stored. The end goal is that the user need not know where data are stored but will be able to view a list of all of them, organised in some folder structure. Ultimately, we expect that this will evolve into something like the EBI's repository for published data, so a user might publish an article in a journal along with a URL to where the data might be located. That URL would be location independent so that the data could be found long after the original server hosting it has passed to silicon heaven.

Data Centre

This component will present a standard interface to all access to a resource centre (which would be a better name for it). Ultimately a piece of software may get a handle to access a dataset or service directly but data centre may choose to channel access to all its resources through such an interface. In the AstroGrid VO, this will be the case as we expect most data centres will initially want this sort of monitoring and control. Initially, this component will simply allow access to a dataset. Later additions will allow for data to be routed elsewhere and for data policies to be implemented (checking against community user or group permissions). Finally, some warehousing and intensive data mining facilities will be provided.

Dataset Access

This component will simply take a standard query from the data centre, translate it into the form appropriate for the dataset and execute it, returning the results to the data centre. This simple functionality will be provided in the early iterations. At a much later stage, the ability to create a warehouse for storing query results will be added and the ability to run the user's own code to extract results.

Visualisation

Owing to time/resource constraints, we will not be undertaking research into server−based, interactive visualisation as originally intended. We do intend to provide the option for sending data to a tool which generates an image from a set of data and loads it into the appropriate part of a web page. We will also provide links with one or two desktop visualisation tools and will publish documents which allow any other tool provider to adapt their tool similarly.

Astronomical Tools

We will VO−enable a small number of existing astronomical tools, those considered most essential to operations which fulfill our key science goals, so that they may be added into a workflow or executed directly from the portal.

3 SCIENCE DRIVERS FOR ASTROGRID−2

We are building an infrastructure that will improve and accelerate astronomical research; make comparison and integration of data seamless and transparent; and make manipulation and analysis of datasets as easy and powerful as possible. AstroGrid−2 will go further, for example by providing improved workflow capabilities, e.g. in the development of ontological processes which involve directed workflows for common sets of tasks. All these goals are quite general, but in order to build a system that meets astronomers needs we have analysed concrete example science cases. We describe below how this has been done, and how the process will be repeated and improved for AstroGrid−2, with a special new focus on *personalisation* and capturing direct user requirements for new tools that work with the infrastructure. As well as focusing on specific science cases, we also have in mind preparing UK astronomers for specific future missions and facilities that are of key importance. Finally, this section also notes how the science drivers for the AstroGrid−2 project are aligned with the development of the Euro−VO initiative (of which AstroGrid is a primary member).

3.1 Science Requirements Analysis for AstroGrid

The current AstroGrid project recognised that the creation of a supporting infrastructure to enable astronomical scientific endeavour must be guided and informed by the scientific priorities of the community. This ensures that the capabilities that AstroGrid create are able to meet the demands of the astronomy end users.

To this end, AstroGrid, in its Phase−A period, undertook a rigorous science requirements capture process, led by the AstroGrid Project Scientist.

- Scientific use cases were taken from astronomers associated with the consortium, and from the community at large.
- These were prioritised, and a key set of ten requirements chosen to be representative of both topical science, and processing and analysis needs. (These key AstroGrid science drivers are fully described at http://wiki.astrogrid.org/bin/view/Astrogrid/ScienceProblems).
- The key science drivers were decomposed into use cases. From these the functionality required is being developed within the context of AstroGrid.

3.2 Enhancing support and access to current datasets

AstroGrid−2 will enable continued and enhanced opportunities in using existing and legacy data sources. The vital VO−enablement of existing archival datasets held in the UK will provide continued and enhanced value for the UK astronomical community, especially in the era of the new projects described below. The requirements and proposed activities in this area are described in Section−7). In the near future it is clear that the VO−enablement of existing data collections is likely to be a very high priority for the UK community. Following on from the initial pilot VO−enablement activities in the current AstroGrid, new science areas here will include study of galaxy properties via access to multi−object spectral surveys such as those of 2dF and the SDSS. Such studies have enormous potential for astrophysics research leading to both new understanding from multi−wavelength statistical samples and to serendipitous discoveries that can play a pivotal role in leading to new paradigms. All of this depends on the timely VO−enablement of existing data sets (see also Data Centre activities in Appendix A.1.).

3.3 New Science Priorities

A major initial driver for the AstroGrid project was to support the effective scientific exploitation of major projects emerging in the 2003−2005 time frame, with particular emphasis on the VISTA project (http://www.vista.ac.uk). When it comes into operation in 2006, VISTA is set to provide the UK and European communities with the deepest large area survey of the sky in the Infra−Red. Its precursor, WFCAM on UKIRT, will provide a powerful complementary Infra−Red survey of the Northern sky. The current AstroGrid project will provide the capabilities to deliver the processed data WFCAM data as provided by the VISTA Data Flow System pipeline project, to the community.

The capabilities now planned for AstroGrid−2 aim to meet the demands of other projects identified by PPARC's Science Committee 2003 Strategy Paper and PPARC Road Map as being of high scientific importance to the UK community. These include, in approximate time order:

- exploitation of the UK's investment in European Southern Observatory (ESO) (now)
- the two Gemini telescopes (now)
- XMM−Newton (now)
- SWIFT (late 2003)
- WFCAM (2004)
- VISTA (2006)
- e−MERLIN(2007)
- Herschel (2007)
- Planck (2007)
- Eddington (2008)
- ALMA (2008−10)
- GAIA (2010−12)
- JWST (>2010)

- Xeus (>2012)

The above projects will all produce high value scientific data, injected into the VO by pipeline processing. AstroGrid−2 will provide the broad capabilities needed to maximise the utilisation of these data, with emphasis on interoperability, ease of handling large data flow, access irrespective of end user location and so forth.

By way of illustration of the key new demands made by science programmes exploiting data from these projects, typical science from two of the above is given:

Illustrative Science Drivers from Eddington

Eddington will measure the brightness of stars to high accuracy and will address two key issues:

- the detection of exo−planets by looking for photometric variability in hosts stars due to planet transits
- investigation of stellar structure through astroseismology (again measuring photometric variability in stars to great accuracy

The first of these will require the use of AstroGrid−2 to enable access to the wealth of multiwavelength data collected for the Eddington targets and candidate exo−planets. VO technologies will also support the discovery of transient objects in the Eddington data, collating a new inventory of, e.g. Near Earth Objects. Multi−dimensional mining of the data, enabled by the AstroGrid Data Exploration Framework (see Section−6.1, may reveal objects of high import, especially in conjunction with other data sets.

Illustrative Science Drivers from Herschel

Herschel, with its suite of far−IR to sub−mm instrumentation will enable new insights in a number of areas:

- Galaxy formation and evolution in early universe and the nature of active galaxies
- Star forming regions and interstellar medium physics in Milky Way and external galaxies
- Molecular chemistry of cometary, planetary and satellite atmospheres in solar system

Science areas like these will require the use of the capabilities of VOs and AstroGrid−2 in particular to enable access to complementary data sets at all wavelengths from radio to X−ray and beyond. Science priorities such as probing protoplanetary discs and star−forming regions, or investigating Ultra Luminous IR Galaxies and testing cosmological theories against the properties of high−redshift objects, involve regions which are heavily obscured by dust or which are most securely and rapidly identified using the broadest spectral energy distribution.

3.4 Cross−Fertilisation with Solar System Research

AstroGrid−2 is aimed at facilitating astronomy. However, new missions will create a new community of astronomers and solar physicists with common interests. AstroGrid−2 will support these science areas from an astronomical perspective, offering solutions developed complementary to those developed in the Solar community.

The UKSSVO project bid aims to pull together the huge variety of datasets, including solar (including helioseismology and remote sensing), planetary, heliospheric, magnetospheric, ionospheric and interplanetary. Both AstroGrid−2 and SSVO will use the infrastructure developed by AstroGrid which will enable ties to be made between these widely differing areas of research. Possible joint research areas include connecting the interior flows on the Sun with astroseismology observations from Eddington and investigating basic plasma processes such as magnetic reconnection on multiple scales: from solar flares, flaring in active galactic nuclei, to magnetospheric reconnection.

3.5 Facilitating New Ways of Working

We expect that new tools will encourage new ways of working, and that the focus of the current AstroGrid on seamless querying and data integration will be an example. AstroGrid−2 will develop further innovative concepts and capabilities :

− Topic specific workspaces, giving access to all data and tools relevant to a certain astrophysical problem.

− Aided work−flows, where the user constructs a personalised data pipeline, using VO components, and in a manner where the system provides sophisticated guidance.

− Methods whereby the outputs of data manipulations can be automatically fed back into the operations of telescopes, both for real−time and ordinary observational programmes.

− Advanced ways of visualising and exploring multi−channel data.

3.6 AstroGrid−2 Science Requirements Process

As for the initial project it is vitally important that the deliverables from AstroGrid−2 are aligned with the scientific needs of the community. To this end a system is required to allow the capture and prioritisation of the user science needs.

It is anticipated that the systems and procedures put into place by AstroGrid will be adopted by AstroGrid−2 and adapted where necessary. The project will initially seek out a wide and demanding set of science drivers which will shape the virtual observatory. It is anticipated that the large collection of use cases gathered during the current AstroGrid project will be used as a basis upon which to build. This activity will led by the AstroGrid Project Scientist, and will be completed in the first few months of the new project.

AstroGrid, during its Phase−B development activities, has constituted the AstroGrid Science Advisory Group to advise the project, through the Project Scientist, on a number of issues:

- scientific utilisation of the AstroGrid Virtual Observatory
- advising the project on ranking of scientific priorities
- technical issues as they affect end users (e.g. user interfaces, functionality provision, access to data products etc).
- deployment of the AstroGrid Virtual Observatory and take up through e.g. the beta tester programme

It is proposed that this group will continue for AstroGrid−2. Adjustments to the composition of the group will reflect the updated priorities for AstroGrid−2. For example members representative of the other major funded projects (potential projects described in Section−8) will supplement the AG−SAG, to ensure that the scientific demands of these projects (which will rely on the AstroGrid−2 infrastructure to a greater or lesser extent) are met. The fit of the AG−SAG within the overall project management structure is shown in Section−9.1.

3.7 Personalisation in AstroGrid−2

AstroGrid−2 recognises that new requirements will emerge through the project as the user community develop new ideas whilst working in the new virtual observatory environment. A key concept of AstroGrid is to allow user customisation − for instance, configuration of the AstroGrid portal to reflect their priorities, and their own working areas with MySpace. AstroGrid−2 will take this a step further by developing specific applications and tools to meet a specific user demand (as prioritised via a call, and independent peer review by the AstroGrid Science Advisory Group). This will be managed within the project iteration cycle and will allow development of tools in a 3 or 6 month period. Thus this would allow for instance the grid−enablement of a user code or application to perform an advanced analysis step (e.g. a classifier of objects, SN, extreme Pop III stars, etc) based on a combination of indicators from various data sources. This concept is outlined in Section 4.3, 'Personalising the Virtual Observatory'.

3.7 Personalisation in AstroGrid−2

4. EXTENDING THE VIRTUAL OBSERVATORY INFRASTRUCTURE

The AstroGrid project has defined and will develop a robust and extensible infrastructure upon which the Virtual Observatory will be enabled. In AstroGrid−2 we will extend the infrastructure further, providing important new capabilities. We will build new science tools and adapt existing ones to provide the astronomer with a rich and effective interface to the VO. A programme will be offered to UK astronomers to nominate the development of tools that will directly benefit their own research work.

AstroGrid−2 will also incorporate the investigation into new technologies and techniques designed to make the VO infrastructure and the tools built upon it more powerful, usable and extensible. As these investigations bear fruit, the prototypes they develop will be incorporated into the VO infrastructure.

4.1 Maintenance, Refactoring and IVOA Cover

The AstroGrid project will complete in December 2004 but is providing software on a quarterly release cycle. As beta testers and trial users put the software through its paces, errors (arising from bugs or failure to meet specific needs) will be detected and relayed back to the development teams. AstroGrid−2 will continue to maintain previous releases of software, either with immediate fixes or incorporating changes into new releases of components.

AstroGrid involves the creation of custom technologies in areas that will later be covered by standards bodies like W3C and GGF, and associated projects like OGSA−DAI. AstroGrid will also be ahead of many of the standards finalised by the IVOA. Therefore, during AstroGrid−2, it will be necessary to refactor some AstroGrid components to ensure standards compliance and interoperability with components from other VO projects.

Examples of areas where AstroGrid−2 expects to refactor are:

- database handling in the light of successors to OGSA−DAI
- security handling, in the light of work by the OGSA−SEC WG of GGF
- binary−data handling in the light of the DFDL work in GGF
- compliance with IVOA standards (registry interfaces, item naming, data and image access etc)

In the current context, *refactoring* implies three main activities.

1. Changing the structure of the existing code while preserving its functions and behaviour.
2. Changing the detailed behaviour of the existing code, in order to conform to standards, while preserving the gross characteristics seen by the user.
3. Replacing parts of the code with standard components provided by other open source projects.

As with the AstroGrid project, we will be closely involved in the development of standards and demonstration projects with the AVO (and its successor, Euro−VO) and US−VO projects and the IVOA standards−setting body. This involvement often takes developers away from the work they are performing. It is not possible to assign just one person to these liaison and joint development activities so we will provide cover for them with additional resourcing.

We estimate that, for AstroGrid−2, an overhead of two FTE over the lifetime of the project will be expended on maintenance, refactoring and IVOA cover.

4.2 New Components

There is no doubt that new infrastructure components will be required for the AstroGrid VO. We have listed here two such components which we had insufficient time to develop in AstroGrid. Towards the end of the AstroGrid project, we will review the use of the infrastructure and develop plans for these new components. We estimate a need for two FTE over three years from January 2005 for this work.

4.2.1 Job and Query Estimation

In deciding whether a job is worth running, an astronomer will require information about the estimated duration of the job. Many factors will influence this information:

- duration of any embedded data queries
- input and output volumes
- parameters set within each component of the job

and the information must be gathered from many areas:

- past runs with similar parameters
- 'test' runs on known volumes
- predicted machine loads at the time the job is run

Many projects in the Grid world are investigating the problems of estimation at the moment. We will make use of the information and software from these projects to provide dynamic job and query estimation for VO astronomers.

4.2.2 Extending Access to the VO

AstroGrid−2 will deliver three separate portals into the VO:

- ***Purely server−based and web−delivered:***
  This is the interface that AstroGrid is now developing. It will make the minimum demands on the user's computer. The user will be identifiable no matter where he/she logs in (using username/password or server−side certificate forwarded by proxy).

- ***Client−based portal using server−side back−end components:***
  The main aim of this approach is to provide the user with feature−rich interfaces. Workflow is one component that will benefit from such an approach, allowing us to develop component toolbars, drag−and−drop addition of components into a workflow etc (think of ***Visio for the VO***). The key aspect of this approach is to use the back−end power of existing server−based components through rich interfaces. The user's machine will act as a node on the grid, interacting with other nodes and components on the grid.

- ***Stand−alone client−based portal***:
  This approach will allow the user to work completely offline, only periodically connecting to the internet to upload a workflow or download results. We will need to investigate how a grid node can operate in mainly offline mode and will develop new technologies and standards to this effect.

To achieve the above, some of the AstroGrid components may need to be refactored to ensure that future developments of core code can continue alongside parallel but separate interface developments.

4.3 VO Personalisation

AstroGrid−2 will offer a programme of direct user support by delivering highly focussed tools which exploit the capabilities of the VO. This concept will build on ideas developed in the AstroVirtel system. In this programme, calls were issued to exploit the databases provided by ESO: the successful astronomers would visit a small team of developers at ST−ECF located in Munich, define requirements on a tool needed to be developed, and the tool would be delivered to them. Perhaps the most successful such tool developed to date is querator (see http://archive.eso.org/querator/index.html).

In AstroGrid−2, astronomers will be invited to suggest which capabilities, not currently provided by the VO, should be developed. On a quarterly basis, a number of projects will be selected by the AG−SAG and the AG−PM, based on the scientific case made and suitability of the proposed tools for adoption by other VO users. These projects will then be delivered within the AstroGrid−2 development cycle. The proposing astronomer would become part of the workgroup formed to develop their tools, interacting directly with the AG−PS and the developers.

This concept of reacting to the specific demands of the astronomer has also been identified as an important concept in the recent submission to the FP6 programme for funding of the Euro−VO project. The AstroGrid−2 'personalisation' programme would be aligned with the Euro−VO effort in this area if the Euro−VO bid is successful. Further, it will complement it by providing a programme that reacts on shorter timescales (3 monthly rather than yearly) and allow development of applications aimed at the interface of AstroGrid−2.

In order to produce tools for between one and three proposals per 3 month iteration period, an effort line of 2 FTE (developers) will be required for three years from January 2005.

4.4 Science Tools

In addition to the above programme of custom tools development, AstroGrid−2 will VO−enable existing tools from Starlink, AIPS++, IRAF and other suites and will make use of Grid and VO technologies to provide tools which we understand astronomers are keen to see developed.

Examples of the types of tool include:

- GAIA−like, integrated display, photometry, astrometry and polarimetry applications, for 2+ dimensional data, enhanced by being integrated with the Data Exploration Framework, Agent and other AstroGrid infrastructure (GAIA here refers to the Starlink application which is widely used by UK astronomers).
- 1−Dimensional data handling such as spectra, with line parameterisation including position, width, shape as well as red−shift and structure.
- Time−varying data handling, for example image or spectral data, to determine parameters for time varying and periodic data.
- Tools for creating new types of data object, for example spreadsheet−like objects of models, images, spectra and tables to investigate the impact of changes in one component on others
- Agent−based tool for advising when data appears on a particular object or group of objects.

Some of these software components will simply require the wrapping of existing tools in VO−like interfaces while others will be newly created. Many will require matching to the common data model and standard interfaces emerging from the efforts of IVOA participants.

This is a key illustration of the benefit of integrating tools development with core infrastructure work: tools can be developed and adapted at the same time and by the same people as infrastructure development leading to considerable cost savings and benefits in terms of the spread of expertise across a number of institutes.

We estimate that the above will require one FTE for the full length of the AstroGrid−2 proposal (3.75 years) and another one FTE for three years beginning January 2005.

4.5 Incorporating R&D Prototypes

AstroGrid−2 includes several semi−autonomous sub−projects which will investigate the integration of leading edge technologies into the Virtual Observatory (section 6, 'DEVELOPING NEW TECHNOLOGIES'). These projects will only develop software to prototype level. Once the prototype has proven itself reliable and stable, the technology will be integrated into the VO by developing new components and modifying existing components of the AstroGrid infrastructure.

This integration will happen as a normal part of the AstroGrid−2 development cycle. In this way, disruption to the VO is avoided, new technologies are properly integrated with existing ones and cost savings are realised through resource sharing.

Based on estimates from those developing the R&D sub−project proposals, we estimate an overall requirement for three FTEs over three years beginning January 2005.

4.6 Helpdesk and Training

The introduction of radically new software and new ways of handling known datasets can be very confusing. AstroGrid−2 will initiate a training programme and set up a helpdesk located at one of the key DCA Data Centres (see section 7) with staff on hand to answer queries about using the AstroGrid and AstroGrid−2 software.

The trainer will:

- visit astronomy centres to train users in AstroGrid and AstroGrid−2 software
- visit data centres to train staff in installing and maintaining VO infrastructure
- create online training programmes

We will also install helpdesk software which will provide an online facility to:

- allow users to file questions and error reports
- look up already answered questions and logged errors
- allow helpdesk staff to populate an online knowledge base
- create and maintain a 'FAQ' page

The software for the training programmes and help desk will be chosen by the AstroGrid Technical Support Panel from among many open source and commercial packages available (the investigation will be funded and carried out within AstroGrid, estimate is approx 1 sm).

We estimate that the staffed help desk will only be required for a period of twelve months from April 2004 with one FTE for that period, and the trainer will be required for two years from April 2004 with one FTE.

4.5 Incorporating R&D Prototypes

5. ASTROGRID IN THE IVOA

5.1 IVOA Standards

The UK's Virtual Observatory does not exist in isolation. A key requirement of any VO system is that it is interoperable with those emerging across the globe. Through its involvement in the International Virtual Observatory Alliance (IVOA), of which it is a founding member, AstroGrid−2 will ensure that all the software developed is compliant with the emerging standards and will be able to influence the setting of those standards. There are a number of working groups in which AstroGrid personnel are already involved.

5.1.1 Resource Registry

This working group is led by the AstroGrid Project Manager and its role is to ensure that the registries being set up by different VO projects will allow an astronomer to be able to locate, get details of, and make use of, any resource located anywhere in the IVO space, ie in any Virtual Observatory. The IVOA will define the protocols and standards whereby different registry services are able to interoperate and thereby realise this goal. In AstroGrid−2, radical new technologies and techniques will be developed to assist astronomers in their work and we will work to ensure that these resources can be accessed from the VO in as simple a way as datasets will be from the current round of standards setting.

5.1.2 Data Modelling

This group will define standard ways in which data items are referred to leading to more consistent data representation in the VO world. As standards emerge, we will incorporate them into our systems. Again, in AstroGrid−2, the creations of new types of resources and new ways of referring, cross−referring and representing data will feed back into this standards process.

5.1.3 Content Description (UCDs)

The Unified Content Descriptors (UCDs) developed at CDS are a *de facto* standard for describing the content of data columns (though emphatically not a replacement for column names). As such, they are a very first step in the process of developing an astronomical ontology. The AstroOntology effort will work closely with CDS and this working group to ensure that ontologies are backwards compatible with UCDs so that software can make use of the greater complexity of ontologies only when it becomes suitable.

5.1.4 Data Access

The task of this working group is to define and formulate VO standards for remote data access. Client data analysis software will use these services to access data via the VO framework; data providers will implement these services to publish data to the VO. We expect that much of this work will be complete as AstroGrid−2 begins but new standards will emerge as more powerful resources are added to the VO, particularly those defined in section 6.

5.1.5 Query Language

A standard query language applicable to all datasets and valid across all domains of astronomy has long been the 'holy grail' of astro−development staff. The reality has been, though, that deploying data on different platforms and using different technologies has made this dream a nightmare. However, with the VO and its standardised interfaces for accessing data, it is now possible to implement a standard VO query language. AstroGrid is already implementing such a VOQL based on an xml−ised form of SQL extended with specialist astronomical functions and based on designs currently being considered by this working group. Later work in this area will incorporate ontology efforts within AstroGrid−2 (section 6.4) and other VO projects.

5.1.6 VOTable

VOTable was the first standard accepted by the IVOA and developed by the loose grouping of researchers and developers which preceded the formation of the IVOA. It is a standard for the exchange of astronomical data and is rapidly being incorporated into many existing astronomical tools. Other tools are being developed which enable the visualisation and manipulation of data in VOTable format. Work is now beginning in this group to incorporate other types of astronomical data into the VOTable standard. AstroGrid−2 will play its part in this process.

5.2 Demonstrations

The IVOA is keen to demonstrate the usefulness of the coordination of its members' projects and thus will organise periodic demonstrations of interoperable VO software. AstroGrid−2 recognises the benefits to the VO enterprise as a whole from these demonstrations, not just in that they bring the work we are doing to the notice of a wide group of astronomers but also because it concentrates effort on ensuring interoperability among systems. We will contribute to these demonstrations with suitable resourcing.

6. DEVELOPING NEW TECHNOLOGIES

6.1 AstroGrid Data Exploration Framework

6.1.1 Introduction

The success of the VO depends on the ease with which science can be extracted from the data it holds, most importantly through the new kinds of analysis that the VO makes practically possible for the first time, such as the exploration of the multi−dimensional data spaces created by the federation of distributed databases. AG1's data exploration work was curtailed due to restricted funding, so AG2 must develops the framework for doing this new science. This Section describes plans for providing the UK astronomical community with a fully−functioning data exploration capability, featuring a suite of basic algorithms, as well as, crucially, an extensible framework within which they can deploy their own data exploration tools.

6.1.2 An ADEF Usage Scenario

An astronomer has a hunch about connections between the properties of brightest cluster galaxies (BCGs) and those of their host clusters. She queries the VO to construct a sample of clusters and BCGs observed with sufficient detail, yielding a set of 400 attributes for 10,000 BCG/cluster pairs. The astronomer runs a statistics package which seeks the twenty attributes with the highest information content, and then generates a grid of scatter plots for pairs of them, arranged in order by the strength of the correlation between them. This reveals that there are very significant correlations between a set of six attributes, so the astronomer launches another visualization tool, which allows navigation through projections of a multi−dimensional data space. The astronomer selects a subsample of 200 objects to visualize, chosen by a statistical algorithm to be representative of the distribution of the full sample across the six attributes of interest. This reveals three clusters of points in this data space, presumably corresponding to distinct populations, which the astronomer defines as three classes. This classification scheme is then applied to the full set of 10,000 records, and statistical tests run to assess its significance. This is found to be strong, so the astronomer saves the data from this session, and moves on to figuring out the astrophysical processes that might lie behind this division into three classes.

6.1.3 Issues to be solved before this scenario can become a reality

Association Techniques: A key component of the VO is the ability to associate entries in different databases − e.g. an X−ray source with its optical counterpart. Web services exist for performing simple matching by proximity, this is not adequate in the general case. Work is thus required on implementing more complex association routines and this should be integrated with the Distributed Query Processor being developed by the OGSA−DAI/DAIT project to provide a VO query service that includes the making of associations. Default associations between popular pairs of very large databases (e.g. UKIDSS and SDSS) should be stored, rather than computed repeatedly and expensively on−the−fly, but it is not clear how or where that information should be recorded. Within the biological community, for example, it is commonplace for interested scientists to provide annotations to published databases and these are incorporated within the database query mechanism in such systems as the Distributed Annotation System (DAS). We propose to develop a *Distributed Association Service*, mediating the identification of multi−wavelength IDs in the VO in a similar manner.

ADE−optimised partial mirrors of datasets: Many data−intensive operations commonly used in astronomy can be executed much faster if the data are stored in column order, rather than in row order, as is conventional. Commercial packages exist for this (e.g. Sybase−IQ), but these appear prohibitively expensive for general deployment within the VO. Within AG1 we shall assess the performance benefits from creating partial mirrors of the most popular datasets (e.g. positions, magnitudes and star/galaxy classifications for all SDSS and UKIDSS sources), optimised for ADEF operations and stored in, say, column−oriented FITS tables or a vertically−partitioned, compressed XML format. AG1 is involved in initial experiments on both of these techniques, and these should mature in time to be deployed in AG2.

Alternatives to verbose XML: The multitude of data formats used by different data sources and analysis tools as a major hurdle to data exploration within e−Science. The likely solution to this problem is the use of XML data formats, such as VOTable, and the provision of XSLT transformation tools. The verbosity of XML makes it inefficient for large data volumes, but computer science research suggest possible solutions to this problem: XML documents can be restructured to facilitate much more efficient querying or can be transformed into a binary format when compactness is crucial; an example of the second approach being the development by the eDIKT project at NeSC of BinX, an XML schema for describing binary data. AG1 is engaged in tests of both these technologies, which should be maturing by the time that AG2 starts, so that they can be implemented in ADEF.

Visualization of multivariate data: Perhaps the two most successful techniques for the visualization of multivariate data are the use of grids of 2D scatter plots, which aid the identification of correlations between pairs of attributes, and parallel co−ordinates, where each record appears as a "polyline" linking a set of parallel axes representing the attributes, and within which clusters of nearly identical polylines can be seen as indicating similar records. Both methods suffer from scalability problems, so recent research has sought to combine an initial pre−processing step, whereby the number of attributes and records is reduced by cluster analysis, for example, and the resulting visualization is less complex. The work of Ward is especially important, and is made publicly available through the Xmdvtool software, which we shall use to develop a prototype system for visualizing multivariate astronomical data. We believe that Xmdvtool will provide a sound platform of multivariate visualization software, which we shall enhance in a number of important ways, tailored to the AstroGrid community. There are a number of other free data visualization packages available (e.g. Partiview, OpenDX, Paraview, SciGraphica, DataPlot, Grace, VOPlot and GNUplot), so we shall also survey these, to see which are best suited to the needs of AstroGrid and easiest to wrap into a web/Grid services framework.

Data Mining: There exits a vast array of data mining software (see e.g. KDnuggets), but data mining in the commercial world is often very different from what is required in astronomy. So, we shall identify which classes of data mining task are relevant to the VO − clustering and classification algorithms, data cleaning, quality checking, regressions, multivariate analysis, principal component analysis, Fourier and wavelet transforms, Voronoi tessellation, and self−organising maps, plus statistical functionality, to carry out standard statistical tests, find outliers, detect signals in noisy data, form randomly sampled subset of large datasets, etc − and then assess which of the existing software to implement them can simply be wrapped appropriately for working in a web/grid service environment. Some will have to be re−engineered to meet the requirements that the VO has regarding scalability (both in terms of the number of dimensions that they can study and the data volume, since many are designed to work in main memory) and the use of distributed data sources. Where simple wrapping is not possible, and a particular algorithm is felt to be necessary for ADEF, two choices exist: to recast the algorithm (most likely so that it processes records serially, or by chunks, rather than requiring all the data in memory at once), or simply to ensure that large memory machines are available to run such jobs, and ADEF will assess the potential trade−offs between these options.

Composing data exploration services: Another major issue is how to compose different tools within a web or Grid service framework into a workflow and there is already work on this underway within the e−Science community: for example, the DiscoveryNet project is producing a prototype architecture for composing data mining services within a Grid environment. Many astronomers may feel happier using Interactive Data Language (IDL), or Starlink routines, rather than learning new tools, so we must ensure that ADEF's extensible service framework can wrap familiar packages and readily allow the integration of new ones, developed by users. A particularly important example of composing data exploration services is the requirement for tightly coupling data mining and visualization services to achieve the functionality described in the usage scenario of Section 6.1.2. In that case, it would be necessary to integrate dimensional reduction techniques (e.g. clustering, principal components analysis, Local Linear Embedding, Isomap, etc) with the multivariate visualization software, so that the dimensionality of the dataset to be visualized could be reduced to a more reasonable level. Similarly, there must be a mechanism whereby labeling of data points in a visualization tool can define a subset upon which to run a data mining algorithm.

6.1.4 Milestones and Deliverables

Date	***Milestone***	***Metric***
Jun−2005	Science Requirements Capture	Document delivered
Dec−2005	Xmdvtool demo	Users can visualize data from VOTable
	Data mining testbed	Users can run testbed algorithms from AG portal
Jun−2006	ADEF−optimised partial mirrors of datasets	Testbed services run on mirrors
Dec−2006	Associations in Distributed Query Processor	Use can make associations as part of query
Jun−2007	Deployment of ADEF hardware	Users can run testbeds on hardware
	Integration of data mining & visualization services	Scenario in Section 6.1.2 is possible
Dec−2007	Test ADEF tools	Working ADEF

Date	***Deliverable***
Jun−2005	Documents: Science Requirements Analysis; survey of existing visualization tools
Dec−2005	Documents: design of Distributed Association Service; visualization functional requirements; survey of existing data mining tools; hardware design for ADEF
	Services: Xmdvtool demo; testbed data mining services
Jun−2006	Documents: design incorporating association techniques into Distributed Query Processor (DQP); architectural design for visualization services
	Resource: ADEF−optimised partial mirrors of popular datasets
Dec−2006	Document: design of core data mining services
	Services: prototype visualization services; incorporation of association techniques in DQP
Jun−2007	Documents: design for collaborative visualization; deployment of ADEF hardware; integrated data mining and visualization services
Dec−2007	Documents: design of ADEF extensible framework; user and developer guides to ADEF
	Services: collaborative visualization services; Distributed Association Service; working ADEF

6.1.5 ADEF resource estimates

We estimate that the plan of work described above will require a team of three researchers and three developers. One of the latter will be supplied by the eDIKT project, so we are seeking five FTE from AG2: three postdocs (Bob Mann, Clive Page, and one other) for R&D and two FTE of effort from AG2's pool of developers. It is clear that it will be the data mining side of the system that will be computationally demanding; the visualization side may even be sufficiently lightweight to be performed at the client end, rather than on the server side. We currently believe that ADEF requires three moderate−sized computational clusters, located at data centres, with 10TB of disk space in total, and we estimate this to cost GBP 100k.

6.2 Intelligent Agents: Building on eSTAR

6.2.1 The role of intelligent agents.

Conventional data mining and analysis techniques are confined to answering the question which the astronomer asks. A more powerful approach is to try to answer the question that the astronomer might have asked if they had been better informed. This is the task of intelligent agents. Their role can be illustrated using an analogy with travel agents. If you have a very specific travel need you can simply and easily fulfill this need by direct booking. However, as soon as your travel problem becomes ill−defined, and your needs more complex, your itinerary must be correlated with available transport, flight and train changes and hotel availability. This process is carried out by a travel agent, who will use their expertise, and knowledge of resources and processes to sift the data and try out solutions. They will then present you with only the information which will be interest, sometimes finding classes of solution you had not considered. In a similar way, though the analogy is not precise, an intelligent agent is a piece of software which can take rather vague requests, obtain data pertinent to the problem and, make decisions based on those data. It can then make further data requests and high level scientific analysis, before finally returning results to the astronomer.

6.2.2 Progress to date.

This work in this section is proposed by the eSTAR project (www.estar.org.uk), who are the leading group in the world working on intelligent agents for astronomy. This proposal will build on this lead to place UK astronomers in a position to dominate the exploitation of the emerging VO. Our work in this area began with intelligent agents which carried out data mining operations to allow them to interpret observations from telescopes, and plan future observations. This proposal is purely for data mining work, grid−enabling telescopes forms a separate proposal outside AstroGrid (see Section 9.5).

The first system was an e−Science demonstrator funded by the DTI/EPSRC core programme. It shows how an intelligent system can react to finding dwarf novae in their bright state (e.g. Allan et al. 2003, Naylor et al. 2003). First, an agent requests telescopes to make observations of many fields containing dwarf novae. By comparing the new observations with archive data the agent works out which stars in a given field have brightened. For each such star it queries SIMBAD to find out if there is a known dwarf nova at that position. If there is, follow−up observations are requested.

The system we are currently working on is designed to address a real science programme, and is funded through the PPARC e−Science initiative. The aim is to obtain prompt spectroscopy of a Gamma−ray burst. On receiving the alert, imaging observations will be requested from UKIRT (by adding an observing block into the UKIRT observing queue). Once obtained, the data will be reduced in near real time, and cross−correlated with 2MASS, to locate any large amplitude variables. After checking with SIMBAD that any variability is outside the known range for this star, it can be identified as the Gamma−ray burster. We will then place a request for a followup spectroscopic observations into the UKIRT observing queue.

Both these systems are far ahead of anything created by other groups, but are clearly experimental systems. We wish to use the experience gained writing these to create robust, adaptable systems which astronomers can use for their own data mining.

6.2.3 The nature of the work to be done.

The original idea behind this programme was to have intelligent software controlling observing programmes on a network of robotic telescopes. Cross discipline discussions with the computer science community soon made it clear that the ``intelligent part'' of the software was an intelligent agent. At first our model was that these agents, each of which controlled a science programme, were interacting with relatively dumb telescopes. However, it has become clear that a more useful paradigm is that of the of the multi−agent contract model. In such a architecture both the software controlling the science programme, and that controlling the telescopes, are thought of as agents. A negotiation is carried between the agents in which each of the telescopes bids to carry out the work, with the science (or user) agent scheduling the work with the telescope agent that promises to return the best result.

However, it is clear that the multi−agent contract model is largely inappropriate for data mining agents. At the current time we are considering two main models which we may be suitable for such an environment. The first is the collaborative agent model, where several agents pool their expertise to solve a problem. What makes the collaborative agent model so attractive is that we can split the science agent itself into several smaller agents, each of which can be provided by grid services.

An example of this model, and possible use case, is that of finding a sample of pre−main−sequence (PMS) stars, over a large area of sky. Faced with such a problem an astronomer should be able to construct an agent which will begin by finding all those objects already classified as PMS stars by querying various resources, such as SIMBAD at CDS. It could then find what information is available about these objects, by mining a range of archives and databases, resulting in a list of known properties. The agent should then correlate these properties, finding where the known PMS stars occur in the available parameter spaces. For instance, the agent will quickly learn that PMS stars occupy particular places in colour−magnitude diagrams, and sit within a particular range of X−ray to optical luminosity ratio. But there may well be other, as yet undiscovered, indicators which the agent could discover on its own by examining the data. To all these indications the agent should then assign a weighting, and then use them to decide on PMS status, producing a final list.

The resulting lists will almost certainly have inconsistencies. We might expect objects which are currently classified as PMS stars to fail the agent's PMS test; or objects which pass most criteria to fail on one crucial criterion. At this stage human interaction could be used as a winnowing process. However, we envisage progressing beyond this stage, to make a connection back to the final science product, the literature itself. The agents we are proposing here would give a list of the most important papers for these outlier (inconsistent) objects, and the papers where the discordant data are discussed.

Perhaps the most crucial agents in the collaboration are those which understand the parameter space; ones that know how to construct, for example, colour−magnitude diagrams. These will use other agents to collect their data, and feed the results from their parameter space exploration, to agents that can characterize the results, perhaps by principal components analysis, or through neural networks. Finally, there will be a decision agent which in the above example uses the learnt process to identify PMS stars, and find the papers for mis−matching objects and data.

The real power of using grid services and the OGSI framework should now be clear. It gives the agents (at little programming cost) access to the best data mining algorithms ADEF can supply (see Section 6.1). This will free developers in this part of the proposal to concentrate on issues beyond the scope of classical n−dimensional data exploration, such as databases of highly interpreted information (e.g. those which classify objects), integration with the literature and decision agents. However, the problem with this model of collaborative agents is that such a framework possibly requires shipping large amounts of data between nodes. Whilst table data are normally small (for instance, our first prototype ran in seconds and interrogated five separate databases in real−time), this is a potential problem. If the agents become bandwidth limited, then we may have to move to a different paradigm. One such is mobile agents. Mobile agents are pieces of smart software that move around the Grid to where they are needed, performing data analysis, reduction and data mining tasks at the most appropriate Grid node.

6.2.4 The Agent Toolkit.

Clearly our first task is to decide on the correct architecture. We will do this by exploring a significant number of science use cases, and then coding two test agents. We have learnt an enormous amount from building our test agents for telescope systems, and intend to follow the same pattern for the data mining systems. Where it is available the test agents will use the emerging VO infrastructure, but as our telescope test agents showed, this is not crucial (since we can access most of the test datasets by other means). In the past these limited task test agents have taken us about six staff months each to build, so we estimate 1 staff year for this activity.

Once the architecture is decided we will build the agent toolkit. This will allow astronomer to build agents from modules using a drag and drop GUI without them needing to understand the underlying computer science. In addition there will be a framework to bind the modules together, and finally ways of creating user modules in a variety of languages. The time for the agent toolkit will be 2.25 staff years for the framework, 2.75 for modules, and 1 for a graphical user interface (GUI) and to add user documentation (including cookbooks).

6.2.5 Linking data and literature.

NASA's Astrophysical Data Service (ADS) are about to begin running optical character recognition software on the papers they have scanned. We already have links with the ADS, for example our Perl modules are now the standard API to ADS, and they have agreed that we can send a programmer to Harvard−Smithsonian Center for Astrophysics (CfA) for an extended visit to work on this dataset (see attached letter of support). This collaboration raises some interesting possibilities. The most obvious is that we should be able to make better ties between the literature and datasets. Observational sections of papers tend to be written in a fairly standard way, and proximity searches between dates and known instrument names should allow us to make a link from a paper to the appropriate database. Using this link in reverse will allow us, in the example outlined above, to go from discordant data to the papers which discuss it (if they exist).

The above is a form of contextual analysis, but in addition other algorithms will be useful. For example, a paper which mentions the name of an object many times is probably important for that object, especially if it is cited by other papers which mention the object frequently. Google (Page & Brin 1998) uses a similar technique to calculate its Page Rank for web pages. This will allow us to find the most important papers for particular objects. Data from highly ranked papers would be given more weight by the agent in its decision making process, and would be drawn to the attention of the agent's astronomer. In our use case above these techniques could be used to direct the astronomer follow−up of objects not classified as PMS stars by the agent, but which are classified in a database as such, i.e. anomalous objects. We allow 1 staff year for algorithm development, and an additional year for module coding.

6.2.6 Total Resource Requirement.

In addition to the normal AstroGrid formula funding we will need funding for one extended (three or four month) visit to the US for the work at the CfA with ADS. After the expiry of the period covered by this application, we also propose employing one programmer on bug fixes and ongoing maintenance (not applied for here). Thus our total staffing requirements are as follows.

Period	***Task***	***Effort***
Apr 2004 − Mar 2005	Prototype Test Agents	1.0*
	Investigate Literature Algorithms	1.0
Apr 2005 − Dec 2005	Toolkit Framework, including testbed modules	0.75* + 0.75 + 0.75

Jan 2006 − Dec 2006	Data Mining Agent Modules	1.0* + 0.75
	ADS Literature Agent Modules	1.0
Jan 2007 − Dec 2007	Data Mining Agent Modules	0.5* + 0.5
	User Documentation + Cookbooks	0.5*
	User GUI	0.5

Named personnel employed on this effort will include Alasdair Allan (1.0 FTE for 3.75 years, effort marked with *), including responsibility for over all architecture and project management. Resourcing will come from personnel employed exclusively on this work and from the pool of developers working on infrastructure (who will take prototypes and integrate them into the core infrastructure).

6.2.7 Tasks, Milestones and Deliverables

From this we can identify the following milestones, and deliverables, by which progress can be measured (assuming a start date of 01−Apr−2004):

Date	***Milestone***	***Deliverable***	***Metric***
Jan−2005	Architecture document completed	Architecture Document	Document delivered
Dec−2005	Agent Toolkit framework work completed	Prototype Agent Toolkit	Toolkit testbed functional
Dec−2006	First agents using toolkit modules + framework functional	Initial Release of Agent Toolkit	Initial release of Agent Toolkit
Dec−2007	Final release of toolkit including GUI interface and user documentation/cookbooks	Production Release of Agent Toolkit	Final release of Agent Toolkit
	Review of sub−project	Grant Report	Final write−up delivered

6.3 Astro Visualisation

6.3.1 Introduction

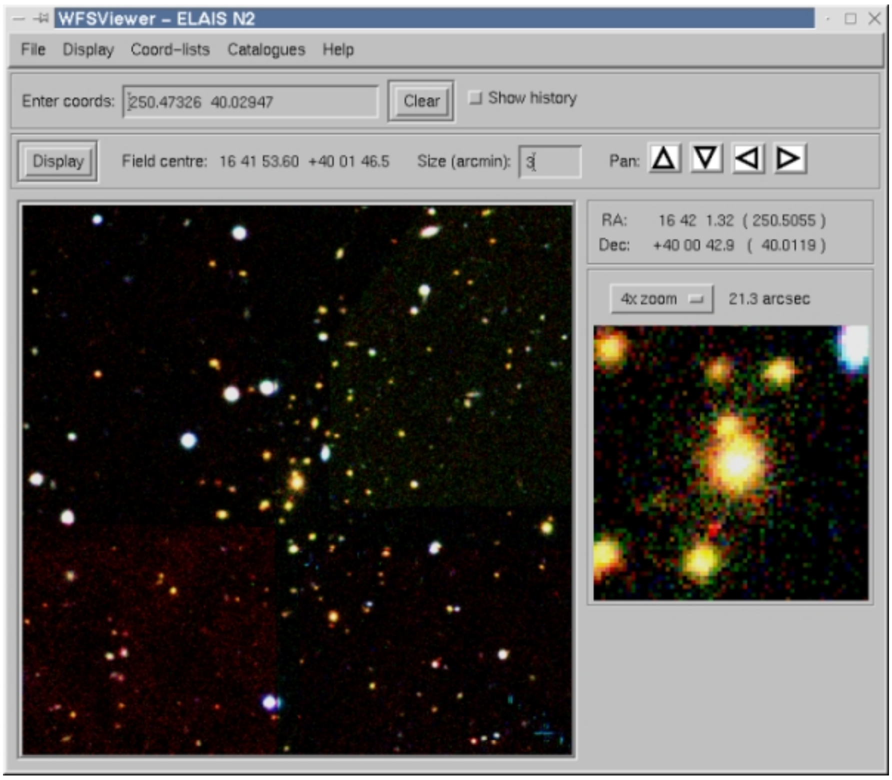
Transparent and scientifically−useful interfaces between users and the Grid/VO are crucial for the success of the virtual observatory. Users must not be forced to juggle multiple applications to evaluate, explore and use the data provided. Consequently data visualisation is key to the entire user experience of the VO. Current data visualisation and analysis tools in astronomy are often tied to format of the data, for example with image display programs tied to single data frames from individual CCDs rather than as a continuous image of the sky where the (real or pseudo) colours represent the dimensionality of the data set. These programs either do not connect to catalogues of other data or do so in a crude way, requiring the user to write their own scripts to correlate different multidimensional data sets.

Much of astronomy will be led by ground−based surveys of large areas of sky, such as the INT, EIS, UKIDDS and VISTA surveys. Few of the existing surveys (e.g., EIS, INT) have resulted in many community publications, despite being carried out for the community. Similarly, the WFI on the ESO/MPI 2.2m telescope has dominated dataflow into the ESO archive for the past several years, but has resulted in few publications. This would be improved by expanding the capabilities of visualisation tools capable of dealing with multi−colour (multi−dimensional) and large area data−sets. Data visualisation in true− or pseudo−colour provides clear and immediately−useful astronomical information, and must be able to cope with the huge number of pixels in an image and the image−to−image boundaries of large area surveys. No existing public visualisation tool can cope with these problems, or is operable in anything other than a rudimentary way with the breadth of access provided by a virtual observatory.

The tools to deliver visualisation of Grid/VO data must be developed in parallel with the infrastructure of the Grid and VO and be compatible with it. Furthermore, the late development of visualisation tools would lead to slow takeup of Grid/VO infrastructure. It is therefore clear that a visualisation tool must be a high priority for the Grid/VO, and so such a development is proposed at this time.

6.3.2 Visualisation Framework

As part of a Leverhulme−funded project to explore how large astronomical surveys (in this case the INT Wide Area Survey) can be best used, the consortium members at Bristol developed a prototype framework for visualising the results of large optical and IR surveys in combination with the results of other surveys in other bands. We developed this framework as there were no useful alternatives available at the beginning of the project (Oct 2000), and have continued to improve it since no alternative has become available subsequently. Our prototype was designed to work as the astronomer does, rather than be limited by the parameters of the data being used. It unifies access to images and catalogues in one interface and presents the images as they are intended to be used: as continuous "true" colour or multidimensional representations of the sky, rather than as individual pointings from telescopes.



Enter coords: 250.47326 40.02947

Field centre: 16 41 53.60 +40 01 46.5

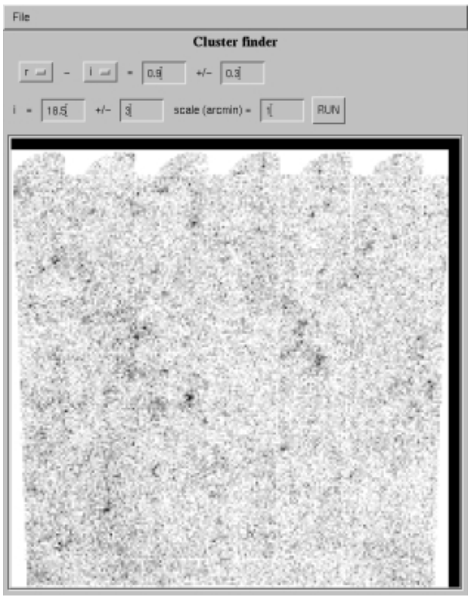
Size (arcmin): 3

RA: 16 42 1.32 (250.5055)

Dec: +40 00 42.9 (40.0119)

4x zoom 21.3 arcsec

View of main screen of Bristol−developed visualiser. True colour image is fully clickable, returning a coordinate. If the coordinate is near an object all catalogue info is displayed. Analysis tools (plugins) hang from the main menu, and interact with main window, coordinates returned from the main window can be used as input to the widget task. Viewer can display arbitrary size images from a few arcsec to many degrees.



r − i = 0.9 +/− 0.3

i = 18.5 +/− 3 scale (arcmin) = 1

Example plugin, which finds clusters based on photometric cuts in optical/IR catalogues. Grey scale gives the density of objects with given characteristics, the darker the region, the higher the density of the objects. Fully clickable, the coordinate returned is transferred to the main viewer so the user can immediately see the true colour image centred at the given coordinate, and then use further plugins to analyse the cluster in detail (eg create colour−magnitude diagrams, generate multi−object spectroscopy masks, return redshifts of objects, thereby determining velocity dispersion of clusters etc).

Interfaces to tasks to perform data−mining and analysis of image data or catalogues can be added to the framework, and the results of these tasks can be overplotted on the data display or in separate windows. Consequently procedures that previously required bespoke scripts and switching between display, analysis and other tasks can now be completed by the user in a few mouse clicks. This greatly speeds user interaction, and minimises the cycle time between analysis, check, and reanalysis of the data. The visual display of data and catalogues has been found to allow quick detection of errors in either (e.g., missing or double−counted objects, objects with incorrect photometry). In short, the framework that has been developed is far in advance of the current standard visualisation tools used by most astronomers in terms of ease of use and potential. When it comes to the visualisation and analysis of large survey data there is currently no other software that can compete. This is largely because the code was written by a research astronomer to facilitate his research, but with a view to providing a more general framework that could be made Grid/VO compliant. Consequently, the prototype framework allows the astronomer to interact with the data as dictated by the needs of the research, rather than the data's internal structure.

The prototype is now in use both at Bristol, where scientific exploitation of the INT WAS survey is being carried out with it, and at other sites in Europe where it is being used to analyse ESO 2.2m/WFI data, with a view to extending it to VST and VISTA.

As it stands, the framework is a functioning prototype and demonstrator for internal consumption at Bristol and its collaborators. It is documented, but not at a level for general support, and its internal structure is not uniformly Grid/VO compatible. Because of its fast−track development, it was coded to be web and network aware but is written as a stand−alone piece of software. It must therefore be regarded as a ***basis*** for an AstroGrid utility and as the means to leverage the knowledge gained in developing the prototype.

In order to integrate the visualisation system into the VO/AG environment we will:

- investigate visualisation tools currently available in the VO context, e.g. Aladin, VOPlot
- carry out a full assessment of the current architecture in order to detect any limitations it imposes on integration
- re−engineer aspects of the design that conflict with the VO/AG structures
- refactor existing code to use the emerging and existing Astrogrid/VO standards (the internal architecture needs to be built around distributed, rather than internal, resources, eg using external databases instead of internal tables)
- bring the code under the development, change−control and documentation regime of AstroGrid
- make the code compliant with the AstroGrid/IVOA software licence

Once the architecture is in place, the plug−in interface will require full specification and development. A library of prototype and general use plugins will be developed to cover the main science drivers of the VO/AG. This will make the visualisation interface immediately useable for detailed scientific work on release and will provide a basis for the users and associated projects (eg SSVO) to develop more specific plugins. The average astronomer should be capable of designing their own plugins given the documentation and infrastructure provided.

6.3.3 Framework Extension

In addition to the above work on VO−enabling the visualisation framework, we will, in parallel, work with researchers from the SSVO and TVO associated projects to ensure that the framework serves as a basis for a wide range of scientific visualisation tasks. We will then add multi−dimensional capabilities to the framework building on developments currently being carried out between AstroGrid members and Australia−VO members (see section 2.4 and http://wiki.astrogrid.org/bin/view/Astrogrid/AngloAustralianGridDemo2003). We will collaborate with developments at CDS with a view to providing a visualisation system combining the range of data manipulation capabilities described here and the breadth of data access provided by Aladin.

6.3.4 Total Resource Requirement

Development will require both a programmer and a scientist. The former is required to carry out the design, specification and coding tasks. The latter is required to ensure that the work is science led by testing each iteration of the system against real science cases, and in particular is required to design and build the plugin library later on in the project. As with the other R&D streams, prototype work will be carried out by a specialised developer with specifications then handed over to the infrastructure development team.

The major tasks envisaged at this stage include:

Half−Year	***Task***	***Effort (per HY)***
HY1	Analyse current architecture for structure clashes with AG/VO requirement	2.0
HY2	Factorise code to be compliant with AG/VO Document code	2.0
HY3	Expand system to interface to other AG/VO technologies Document code Survey science requirements for base−level plugins	2.0
HY4	Expand system with plug−in interface Survey for Plugins Generation of plugin primitives	2.0
HY5	Generation of plugin Library Integration with architecture	2.0

HY6	Generation of plugin library Final user and progammer documentation	2.0

From this we can identify the following milestones by which progress can be measured (assuming a start date of 01−Jan−2005):

Date	***Milestone***	***Metric***
Dec−2005	Refactoring prototype system completed (V0.5)	Release of system
Jun−2006	Extended VO/AG visualisation system completed (V1.0)	Release of system
Oct−2006	V1.0 Programmer documentation completed	Release of documentation
Dec−2006	System with plugin interface completed (V1.5)	Release of system
Dec−2007	Plugin library developed	Release of library
Dec−2007	Final documentation	Release of documentation

And the following deliverables:

Date	***Deliverable***
Dec−2005	AG/VO enabled system with current user functionality (V0.5)
Jun−2006	System with interfaces to other developing AG technology (V1.0)
Oct−2006	V1.0 Programmer−level documentation
Dec−2006	System with fully specified Plug−in interface (V1.5)
Dec−2007	Full visualisation system inc plugin libraries + docs (V2.0)

6.4 AstroOntology

6.4.1 Introduction

The sophisticated data handling packages used by AstroGrid and other VOs require detailed, standardised data descriptions using agreed definitions of associated keywords and concepts. This need is generally recognised; the challenge is to produce a manageable system which is intelligible to humans and machines alike.

Astronomy has various ways of categorizing information and data including its Thesaurus (see http://msowww.anu.edu.au/library/thesaurus/), bibliographic keywords and Unified Content Descriptors (UCDs, see http://cdsweb.u−strasbg.fr/doc/UCD.htx). The services provided by CDS, based on UCDs, are already very widely used

by astronomers and the IVOA is seeking to extend the semantic usefulness of UCDs rather then replace them. This leaves a long way to go before *complex* operations can be automated using such metadata. We need a metadata system which not only describes data, but also supports its interpretation, including links between concepts and conditional or probabilistic definitions.

In computer science terms, the ***ontology*** of a domain or sphere of interest is taken to be the 'explicit formal specifications of the terms in the domain and relations among them' (Gruber 1993). An ontology thus defines a ***common vocabulary*** for researchers in a domain who need to share information. It includes machine−interpretable definitions of basic concepts in the domain and the relations between them. A practical introduction is given by Noy & McGuinness (2001).

The development of an AstroOntology is determined by science drivers such as the need to:

- Handle registry queries which use context−dependent terms (like 'cluster');
- Set up data queries which can automatically find the right data sets to satisfy the query;
- Compare heterogeneous data sets, which may need separate explanatory information;
- Provide assistance to select tools and stored workflows based on 'similar' tasks.

6.4.2 Background: AG1 & UK e−Science

AstroGrid personnel are playing prominent roles in the development of international standards for data models, a VO Query Language and interfaces for a VO Registry of resources, all of which are due to be delivered by the end of 2003.

The UK is well poised to lead in the development of an AstroOntology. We have core developers of ontologies in the BioInformatics world and are leading the GGF Semantic Grid effort (see De Roure 2001 and http://www.semanticgrid.org/). With AstroGrid, we will also have a fully operational grid−based Virtual Observatory, upon which we can deploy prototypes and can incrementally add ontological features to existing components.

6.4.3 Proposal

This R&D sub−project will have five streams:

- Basic research
- Ontology tools
- Ontology development
- Usage prototyping
- Integration

Research

This will primarily consist of gathering information and expertise from other scientific fields (BioInformatics mainly), creating and documenting a methodology for the work and identifying sources of ontological information. This stream will continue throughout the sub−project, in the latter stages by maintaining the contacts previously made.

Tools

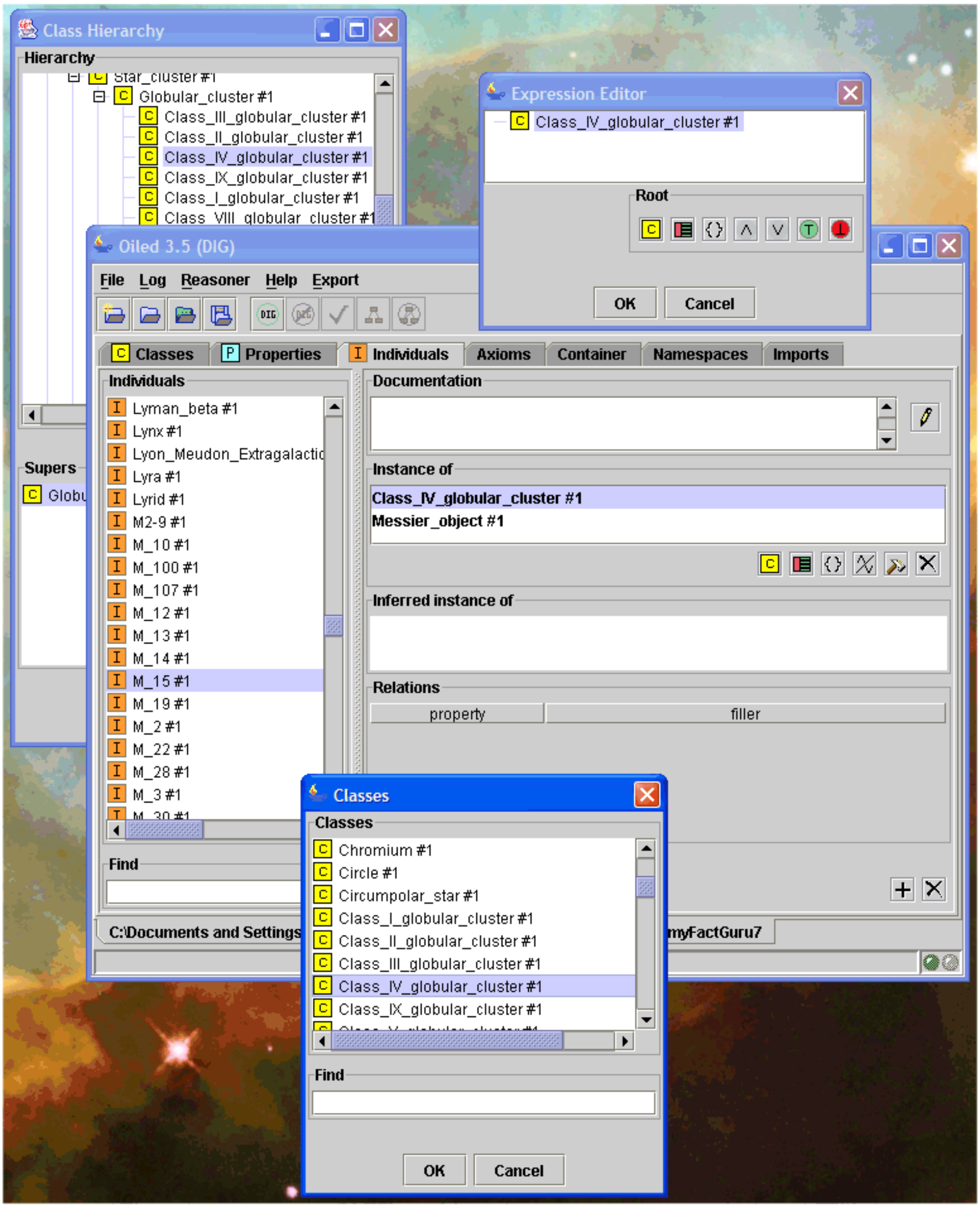
Tools provision will consist mainly of adapting existing tools used in other scientific fields for astronomical usage. We will provide tools which extract ontological information (concepts, relationships) from existing astronomical data holdings and publications. Other tools will be provided to make it easy for astronomers to add to a developing ontology and to annotate existing structures.

Ontology development

The creation of an astronomical ontology will form the core of this effort. This will be carried out by survey, interview and meetings with domain experts and will build on the evolution of astronomical data models within the IVOA.

An ontological approach is ideally suited to encapsulating knowledge in manageable but flexible and interlinked domains so we will set priorities according to science drivers and opportunities for real users to test our work, eg:

- Observational astrophysics in all wavebands, consistent with data models for forthcoming major observatories;
- Astrophysical interpretation, in particular supporting cosmology, galaxy evolution and star/planet formation;
- Solar observations, science and evolutionary models;
- Astro−biology and −chemistry including support data such as molecular and atomic line libraries;
- STP, plasma and solar system science;
- Particle astrophysics.



An example of one tool, OilEd, developed by the University of Manchester, showing a simple ontology extracted from one website: http://www.site.uottawa.ca:4321/astronomy/index.html

Usage prototyping

Once we have a suitable draft of the AstroOntology available, we will begin to prototype ways of using this within the VO, incorporating the applications to science drivers listed above.

Integration

As this effort progresses, we will recommend and specify component changes or new components which the AG2 development team can undertake to implement and integrate into the evolving VO infrastructure.

6.4.4 Tasks, Milestones and Deliverables

The major tasks envisaged at this stage include:

Half−Year	***Stream***	***Task***	***Effort (per HY)***
HY1	Research	Discover existing astro knowledge sources Document work in other fields	0.6
	Tools	Document existing tools Create methodology	0.2
HY2	Research	Contact other researchers Create draft high−level ontology	0.5
		Document how ontologies ***used*** in other fields (relating to VO ideas)	0.2
	Tools	Develop/adapt domain knowledge discovery tools	1.0
	Ontology	Create advisory group; online survey for initial concept list	0.6
HY3	Tools	Continue with knowledge discovery tools Create tools for communal online ontology development	1.0
	Ontology	Create draft ontology Review with advisory group	0.8
	Prototyping	Create & test ontology−based registry prototype	0.3
	Integration	Specify ontology−based registry prototype	0.2
HY4	Ontology	Continue ontology creation & review	0.7
	Prototyping	Create & test ontology−based data query facility	1.4
	Integration	Specify ontology−based data query prototype Specify integration of ontology−based registry facility	0.2
HY5	Ontology	Continue ontology creation & review	0.7
	Prototyping	Create & test ontology−based workflow retrieval	1.4
	Integration	Specify ontology−based workflow retrieval prototype Specify integration of ontology−based data query facility	0.2
HY6	Research	Document results, conclusions & recommendations	0.3
	Ontology	Finalise ontology	0.3
	Integration	Specify integration of ontology−based workflow retrieval	0.2

From this we can identify the following milestones by which progress can be measured (assuming a start date of 01−Jan−2005):

Date	***Milestone***	***Metric***	***Deliverable***
Jun−2005	Existing tools & sources known	Document delivered	Survey of existing tools & sources
	Methodology determined	Document delivered	Methodology for AstroOntology development
Dec−2005	Discovery tools developed	Available and used by researchers	Knowledge discovery tools
Jun−2006	Draft ontology created	Publicly available with tools for navigation	Draft ontology
	Registry prototype	Prototype online and available for testing	Registry prototype
Dec−2006	Registry integration	Ontology−based registry query facility in AstroGrid	
	Data query prototype	Prototype online and available for testing	Data query prototype
Jun−2007	Ontology delivered	Report of review body	Final version of ontology
	Data query integration	Ontology−based data query facility integrated into AstroGrid	
	Workflow retrieval prototype	Prototype online & available for testing	Workflow retrieval prototype
Dec−2007	Data query integration	Ontology−based data query facility from AstroGrid	
	Review of sub−project	Final write−up delivered	Final write−up

6.4.5 Estimates

Named personnel employed on this effort (for the full three years) will include the people so far involved (with their present locations):

- Tony Linde (University of Leicester, AstroGrid Project Manager) (0.2 FTE)
  − to manage the effort and undertake some basic research and specification of prototypes and integration;
- Elizabeth Auden (Mullard Space Science Laboratory, AstroGrid) (0.6 FTE)
  − to undertake basic research, tools and ontology development and prototyping;
- Anita Richards (Jodrell Bank Observatory, AVO researcher)
  − not *employed* on this sub−project but will be involved and provide scientific expertise as necessary.

This effort will also require the services of two PDRAs, each starting about six months into the effort:

- Astronomy PDRA (0.5 FTE for 2 years) − to undertake ontology development
- Computer Science PDRA (1.0 FTE for 2 years) − to undertake tools development and prototyping

6.5 Scripting Environment

6.5.1 Introduction

Many astronomers will be used to using a command line interface (CLI) to interrogate datasets, issue queries etc. It was not possible to undertake the creation of such an environment in AstroGrid − we developed only the simplest web−based, directed interfaces. In AG2 we will want to provide the astronomer with a much more versatile interface which accepts interactive commands and scripted sets of commands.

6.5.2 Proposal

This R&D sub−project will investigate the ideal environment and command structures for an astronomer−friendly CLI to the VO. The scripting environment will make use of existing IVOA standards for interfacing to VO components and will create new standards for scripting interfaces in concert with IVOA partners.

This environment will need to let the astronomer:

- explore and query the registry
- query datasets and other data−oriented resources
- explore MySpace and move data items within the user's areas
- create multi−task jobs and submit them
- check the status of ongoing jobs
- initiate and control VO services

The environment will be available to the astronomer from a web−based interface and as a workstation−based tool (see section 4.1.2: Tri−Portal). In the both modes, the scripting environment will interface to other VO tools. The language developed will be specific to VO tasks but we will investigate the option of layering it on top of an existing scripting language (Perl, Java Beanshell, etc).

As the AstroOntology is developed (see section 6.4: AstroOntology), the scripting developments will integrate the use of concepts and relationships developed within that sub−project and integrated into the VO infrastructure.

6.5.3 Tasks, Milestones and Deliverables

The major tasks envisaged at this stage include:

Half−Year	*Task*	*Effort (per HY)*
HY1	Define needs for scripting environment Specify prototype environment	1.0
HY2	Create simple prototype for registry and data query Specify new VO scripting component	1.0
HY3	Extend prototype to include job construction and control Create infrastructure specs	1.0
HY4	Complete prototype with script creation Integrate aspects of ontology into prototype Document how environment can use ontology developments Create new infrastructure specs	1.0

From this we can identify the following milestones by which progress can be measured (assuming a start date of 01−Jan−2005):

Date	*Milestone*	*Metric*
Jun−2005	Scripting environment requirements documented	Document delivered
Dec−2005	Initial prototype	Users can query registry and datasets
Jun−2006	Extended prototype	Users can create and control jobs from scripting environment

Dec−2006	Final prototype	Users can create full script and execute it, check & control progress

And the following deliverables:

Date	***Deliverable***
Jun−2005	Document: Specification of prototype scripting environment
Dec−2005	Prototype scripting environment
Jun−2006	Extended prototype
Dec−2006	Final write−up of prototypes

6.5.4 Estimates

This effort will require the services of one Computer Science PDRA for a period of two years.

6.6 Continuing Grid Research for AstroGrid−2

6.6.1 Introduction

AstroGrid has relied on new technology both from the commercial/W3C world (SOAP, WSDL, etc) and from the academic/scientific Grid world (Globus, OGSA−DAI etc). Using the former has been a matter of learning, assessing, and engineering, whereas using the latter requires a more intimate involvement with the technology development itself, especially for OGSA−DAI where we have helped to set requirements. Grid technology will continue to evolve, so for AstroGrid−2 we should dedicate effort not simply to incorporating components, but to active *Grid research*.

Grid technology, as typified by the Globus toolkit, tends to operate at a lower level and with a wider scope than AstroGrid products. Any given grid product or standard is likely to solve only part of a problem relevant to AstroGrid; typically, it must be extended or configured to be used for astronomy. Conversely, most grid products and standards are designed to be useful to many fields of science, and most of them have some potential use in the virtual observatory. Furthermore, most developments in grid technology are released and publicized long before they are refined to the point where they can be incorporated in our infrastructure. It can be hard to judge from written descriptions which products of grid technology are useful to us. To judge the true relevance, we need to use the products in prototypes, and these prototypes must be separate from the main infrastructure as they will be unreliable. The prototypes will be disposable code wrapping product code where the latter may become reusable.

6.6.2 Proposal

The research topic will monitor developments in the grid "world", e.g. in GGF, in OGSA−DAI and in the Globus project, and will track the development of new products and standards. When these items reach a sufficient quality, as measured by the disposable prototypes, they will be recommended to the infrastructure−refactoring topic (see Section−4.1.1) for inclusion in AstroGrid−2.

Workers in this topic will need to engage with the grid fora and standards bodies in order to get sufficient information on the new technology. They may also become involved, e.g. through GGF, in setting new standards.

Currently, AstroGrid invests 2 FTE in grid research. (Considerably more effort than this is however planned for *deployment* of mature chosen grid components within the AstroGrid architecture.) From this effort, we cover much of the grid activity and a sub−set of the available products, features and standards. It is hard to be sure how much research effort is appropriate because it is sensitive to the volume, range, and quality of products that emerge from the grid world. Our working plan is to continue to devote 2 FTEs of effort in this direction, but we will revisit the balance of work as the technological scene evolves.

6.6.3 Tasks, Milestones and Deliverables

The major tasks envisaged at this stage include:

Half−Year	***Task***	***Effort (per HY)***
HY1	Define scope for tracking partner technologies Specify study focus	2.0
HY2		2.0
HY3	Create infrastructure specs	2.0
HY4		2.0
HY5		2.0
HY6		2.0

From this we can identify the following milestones by which progress can be measured (assuming a start date of 01−Jan−2005):

Date	***Milestone***	***Metric***
Jun−2005	Assessment of OGSA technologies	Document delivered
Dec−2005		
Jun−2006		
Dec−2007	Final Handover	Project has deployed components based on recommendations

And the following deliverables:

Date	***Deliverable***
Jun−2005	Document: Current appropriateness of OGSA
Dec−2005	
Jun−2006	
Dec−2007	Final write−up and handover of conclusions to VO maintainers

6.6.4 Estimates

This effort will require the services of two researchers for a period of three years.

7. THE DATA CENTRE ALLIANCE: UK E−SCIENCE INFRASTRUCTURE

In order for the AstroGrid developments to turn into a working, successful VO−system, the UK data centres will play an essential role, as they are responsible for some uniquely important datasets, and they will also provide advanced data mining facilities for all UK astronomers. The international context of VO developments also makes it essential for the UK data centres to work more closely together to ensure a coherent service is provided to the UK astronomical community. Such coherence is currently lacking. In the framework of AstroGrid−2 we expect to develop the concept of a 'UK Data Centre Alliance' which will foster much stronger links between UK data centres, acting to coordinate, rationalise and prioritise the VO activities for which the centres will be responsible.

This section describes the essential data centre activities needed: core support for data centres, VO−technology uptake, basic data management together with Grid−node installation and maintenance.

7.1. Core support for data centres

Data centres will play a crucial role in astronomy for many years to come, but this role is becoming increasingly difficult and professionalised. We propose that a small number of UK data centres should be established as "well−found laboratories" so as to be able to deliver the data to the VO as well as continuing to be competitive on the international scene in order to be able to take on new projects.

The existing AstroGrid Consortium already represents the main institutions in the UK which have major astronomical data centre roles:

- Edinburgh & Cambridge: together supporting a wide range of optical and IR datasets and in the near future the UKIDSS and VISTA data;
- Leicester: specialising in high energy astrophysics data from space missions but also holding a comprehensive range of astronomical catalogues in existing systems;
- Jodrell Bank: specialising in radio and interferometer data from MERLIN and other telescopes.

(Appendix A provides a more detailed account of the activities of these centres. MSSL and RAL have equivalent roles in terms of STP data; these areas are addressed in the UKSSVO proposal.)

The strength of these UK data centres lies in their expertise: both with the astronomical data and the instruments used to produce them and with the current systems deployed to deliver them to the community. Core support for these data centres will reinforce their ability to continue with their current data curation roles but more importantly will allow them to evolve into the 'VO Astronomical Data Centres' needed as VO developments mature.

Essentially what this will achieve is the establishment of well−supported and professionally operated UK data centres which will be able to deliver the UK VO as well as being well−placed to take on the data centre role in future projects, placing the UK in a strong position to make compelling bids for these activities. Synergy between different projects will enable effective expertise transfer. Bringing the support of the data centres into the AstroGrid−2 project will also be an important first step in realising the 'UK Data Centre Alliance'.

Core support for these data centres needs to provide for the professional management of centre activities together with adequate system administration activities for the substantial data holdings and software elements.

7.2. Data management

Current datasets held by UK data centres include a large collection which have considerable archival importance. VO−enablement of a significant fraction of these data collections is an important step towards realising the UK VO (see 7.3), but equally there is a clear requirement for the continuing professional curation of these datasets to ensure their continuing availability to the UK community, together with adequate documentation and the maintenance of the body of expertise needed to support effective use of the data by non−expert users.

Those datasets which are purely archival (i.e. are not changing) require relatively small levels of support, but all the UK data centres also support dynamic datasets which are still growing as new data are collected (and/or are reprocessed). Such datasets require higher levels of support. Experience to date has shown that the data management/curation task requires a dedicated staff position at each data centre to cope with the wide range of data handling activities alongside the documentation and user support activities.

7.3. VO−uptake

To some extent it has been implicit in VO developments so far that the task of interfacing existing data collections to the VO infrastructure will be straightforward requiring little additional work. Now that we understand a little better what the VO will actually look like the assumption that the implementation is simple is clearly no longer valid. Connecting existing data collections to the VO systems that AstroGrid is developing will, in reality, be a significant undertaking which will require adequate resources to make it happen. The main areas which will need to be addressed include:

- definition of adequate metadata to describe data collections in ways VO systems can understand, in particular defining UCDs (Unified Content Descriptors) for all columns of tabular data;
- population of existing databases with enhanced or new metadata, especially in respect of data quality and calibration status;
- development and implementation of Web Services interfaces (SOAP and WSDL) that enable VO systems to connect to the wide variety of database systems which provide access to existing data collections;
- migration of data collections to new database systems where existing systems cannot provide the functionality or power needed to support key VO system features or the new access modes which these will enable;
- migration of archival data from obsolescent machine architectures and data formats;
- provision of new database system features such as advanced indexes (based for example upon the R−tree, HTM, or HEALPix) to provide adequate access and search speed;
- provision of Grid Services interfaces based on OGSA−DAI and other OGSI standards, to enable data centre systems to provide computational grid services, where appropriate.

AstroGrid will be undertaking a limited subset of the work involved in its Phase B (e.g. on metadata definition, pilot implementations). Significant new resources will however be required to achieve adequate take−up of AstroGrid developments and thus the VO−enablement of key astronomical datasets. If resources to undertake this work are not made available the UK will find itself in the position of leading VO developments but trailing in the task of actual deployment of data to the VO.

The efforts required to achieve the VO−enablement of existing datasets need to be closely integrated with the data centres that host the data themselves, whilst also forging strong links with the AstroGrid development team so that VO−enablement can commence whilst the infrastructure itself is still developing. These requirements suggest that the most effective way forward is to deploy additional effort at the four UK data centres which already have significant AstroGrid involvement (see 7.1). Flexible deployment of staff resources is envisaged, with allocation to different centres from a small pool of staff.

7.4 Archival datasets held in the UK

The four UK astronomical data centres affiliated to the AstroGrid consortium (at Cambridge, Edinburgh, Jodrell Bank and Leicester) hold a wide range of archival data which remains of considerable importance for UK astronomy. The data holdings include both those for which the UK has primary responsibility, e.g. APM catalogues, and those mirrored in the UK because of their importance for the UK community, e.g. the ROSAT data archive, 6DF.

A full account of the data holdings and the wider activities of these data centres is given in the Appendix to this proposal.

We propose the VO−enablement of ***all*** the significant datasets held in these UK data centres. Given the considerable scope of this task, we expect to prioritise this activity according to, e.g. current usage and science 'scope', in order to ensure timely delivery of the most useful datasets.

7.5 Hardware requirements & Grid−enabling

For the UK data centres fully to take on their role as 'VO Data Centres' will require the provision of new equipment as well as the upgrading of existing equipment (e.g. data holdings storage) and continued support for current equipment. The costs to support existing activities, with a prudent re−equipment plan, can be relatively accurately assessed. Estimating the likely cost of fitting up a data−centre to work in the virtual observatory is less certain as some aspects of the VO's architecture and technological base are still being developed. The additional requirements of equipping a data centre will involve machines to do the following:

- run the service interfaces
- supporting AstroGrid MySpace (serving and managing discs)
- running the community server
- running the WWW portal

(all these tasks are taken from the current Astrogrid architecture which is pretty close to the emergent IVOA architecture). Grid techniques (i.e. OGSI−compliant web services, Globus−style access control and GridFTP) are assumed where appropriate. Additional hardware may be needed if the usage of the local archives increases as a result of the VO.

The required resources are described in the Project Plan.

8. OUTREACH SUPPORT

8.1 Importance of Outreach

Outreach is an important activity for AstroGrid−2. The project needs to inform the professional (and non−professional) community about its activities, but AstroGrid−2 can also provide much more than this by developing tools that can be used in the educational sector. AstroGrid−2 will be in an excellent position to enable PPARC science to be accessed in a friendly and exciting way by the general public. This will achieve several objectives:

- to highlight the functionality of e−Science to the general public
- to enable the general public to analyse PPARC datasets.
- to enable schools to use PPARC datasets for project work. This would expose children both to e−Science and astronomy as an early stage.

8.2 Linking with Outreach Partners

To ensure that the aims of the AstroGrid−2 project is relayed to a variety of sources, the most effective approach is to foster strong links with existing organisations who have a professional outreach role. An incomplete of some of the key partners is given below. For many of these there are existing institutional or personal links with the AstroGrid team.

- Royal Observatory Greenwich/ National Maritime Museum
- The National Schools Observatory
- Planetariums − e.g. Armagh
- Science Centres − e.g. The National Space Centre (Leicester), Explore@Bristol
- Museums − e.g. The Science Museum (London)
- Curriculum enrichment projects − e.g. the Millenium Maths Project
- Outreach activities of other VO groups (e.g. NVO), SSVO
- Schools
- Earth and Space and Astronomy Continuing Professional Development (CPD) programme

Links with industry play a complementary role. Some of the partners with whom we already have contact include research units within IT companies (e.g. IBM (Hursley through OGSA−DAI), Microsoft (BARC and Cambridge), Oracle), but we also need to look to widen these partnerships.

8.3 Planning for Educational Outreach

AstroGrid−2 would aim to form the AstroGrid Outreach Panel (AOP) with membership drawn from the project (e.g. Project Scientist, AstroGrid Outreach developer−coordinator) and members of key external partners such as school teachers etc. The AOP will provide input to requirements for software development for the AstroGrid Outreacher (see below).

8.4 Implementing Educational Outreach

AstroGrid−2 has a unique opportunity to provide tools to enable schools to use and analyse PPARC science datasets in a way that would not have been achievable before the existence of the grid infrastructure. This approach will not only highlight the science, but also the new developments in IT. In order to achieve this we require an AstroGrid−2 'outreacher' whose role would be to develop the software to achieve this and address related issues such as data availability.

We envisage that the AOP would create a list of requirements to achieve accessible and easy to use data for projects at different age groups. For example a 6−8 year old could search for images of stars and galaxies, a 13−16 year old group could search for images of a target in different wavelengths to understand what the differences are between optical, EUV and radio wavelengths etc. A few examples of the type of software that could be developed are listed below;

- School Zone in MySpace: students and teachers can post competition entries, amateur astronomy photos, data reduction projects, long term project data collections (such as tracking solar magnetic field strength over the course of a term), collaborations with other schools.

- Movie Zone and Picture Gallery: simple interface where students can enter a query for movies (solar data movies, artists' impressions of gamma ray bursts) or pictures (planets, galaxies) and see the movie / image displayed immediately. A specialized web service could make this query very efficient.

- Interface: at least two interfaces are required − one with bright colors and easy to click buttons for younger schoolchildren, and a smoother, cleaner interface with real time space and technology news feeds, data query inputs, picture gallery, and access to algorithms for older children or adults.

- Queries: outreach registry queries needs to provide data that requires no further processing.

This kind of project would be a test−bed for future interactions at the school level with PPARC projects.

8. OUTREACH SUPPORT

9. ASSOCIATED E−SCIENCE PROPOSALS

The AG2 proposal comes from a UK−wide consortium, and is meant to provide the basic infrastructural needs of UK astronomical e−Science, including data centre core support, and to lead R&D in key areas, looking towards future technology deployment and a fuller VO data exploration system. AG2 will not provide all the tools that will use the infrastructure, nor will it run pipeline/archive projects that will populate the system. We anticipate strong proposals from a variety of groups that will be completely independent, but which will need to utilise the AstroGrid infrastructure. However there are some projects with which we should have a much closer relationship. In a managerial sense, such projects will be completely independent, but we expect to establish formal links, such as cross−membership of committees, shared documentation systems, and even shared work programmes. Overall, we thereby hope to give UK astronomical e−Science a coherent programme, without creating a overly complex monolithic project.

In this section we detail below (9.1) the 5 other new e−Science projects which are associated with AstroGrid as described above. The remaining subsections describe the other projects and activities with which AstroGrid is linked to give a broader view of AstroGrid's position in e−Science and the VO.

9.1 VEGA: VOProcessing

VEGA is a multi−institution partnership of major groups in the UK active in providing advanced processing capabilities for ground and space based astronomical missions. VEGA has been formed to meet the data processing and and handling challenges presented by major upcoming missions such as VISTA, Eddington and GAIA. VEGA will ensure that the expert knowledge developed during each mission is retained, and made available for exploitation by later more challenging missions, thus allowing the full scientific capability of the programme be delivered. An important motivation for this coordinated approach, in addition to maximising cost−effectiveness and technical expertise, is to position the UK to maximise science return for minimal additional cost. For example, availability of the Vista data reduction system will enhance any UK proposal for involvement in ESO/VISTA survey projects.

The VEGA consortium includes partner institutes also involved in the AstroGrid consortium. VEGA will deliver scientific content to be exposed to the wider community via AstroGrid−2 and the virtual observatory. VEGA will deliver advanced value added products for major missions, and develop scientific focussed access to the processed and derived data and information from the missions. The VO will enable the wider access to these per mission resources.

In order to ensure the seamless interface between the content delivery from VEGA into the VO, there will be technical and scientific cross representation between the projects. For instance, a technical representative from VEGA will join the AstroGrid−2 Technical Support Panel (see AgConsortiumAndPersonnel). Likewise, the AstroGrid−2 project scientist will advise the VEGA project.

9.2 Roadmap To ALMA

The Atacama Large Millimetre Array (ALMA), an array of 64 12−m antennas covering the 30−900 GHz frequency range, is currently under construction in northern Chile. This frequency range bridges the gap between the radio and optical/infra−red regimes and provides a uniquely sensitive probe of the cool neutral material from which planets, stars and galaxies form. ALMA will be 100 times more sensitive than existing sub−millimetre instruments, and is a 'key science driver' in the PPARC astrophysics roadmap.

Although VO compatibility has been specified as a design goal of the ALMA project, this component is neither funded nor scheduled with the project. The development of VO interoperability would be carried out in close collaboration with the ESO ALMA software group and is a natural extension of the work on the ALMA archive currently being done at UMIST and JBO. For ALMA, the archive system is at the centre of all system and science software, and will have to cope with a typical data rate of 1 TB/day.

e−MERLIN, an array of 6 cm−wave telescopes across the UK, will operate at lower frequencies (1.3 − 24 GHz), but will produce a similar volume of data, due to its longer baselines. Although VO interoperability for the existing MERLIN array is being developed as part of AstroGrid and AG2, e−MERLIN represents an increase of more than 2 orders of magnitude in data volume. The e−MERLIN archive will be developed along similar lines to the ALMA archive, will share some of the same infrastructure, and will therefore also benefit from the work proposed here for ALMA.

Interferometers, such as ALMA and e−MERLIN, produce fundamentally different data to direct imaging telescopes. Their basic data product is a set of visibilities, which must be calibrated, Fourier transformed and deconvolved before a recognisable image can be made. There is considerable flexibility in how this is done and users can tailor the process to their own needs including non−image products. This flexibility demands a significantly different approach to VO interoperability for interferometer archives than for many other types of telescope. 3−D datacubes will be commonly produced and access to molecular spectral templates will be needed especially for interpretation of ALMA data. or if short of space just Access to molecular spectral templates will be needed especially for interpretation of ALMA data.

This project requests 3 FTE to develop the Archive Access Layer, control of the reduction and calibration systems and the presentation layer, including appropriate XML descriptors for this type of interferometric data. Real−world testing of the presentation layer will be done using existing MERLIN data and e−MERLIN data which should begin to flow in late 2006.

9.3 VirtU: Theory Side Services

The VirtU proposal (Virtual Universes) brings together significant sections of the computational astronomy community in creating a Theoretical Virtual Observatory (TVO). VirtU will provide solutions to the challenges emerging from mass scale simulations of astrophysical processes. It will enable researchers in their quest in gaining insight from the comparison and confrontation of observational data with these theoretical models.

VirtU will be built upon a dynamic archive constructed, in the first instance, from state−of−the−art cosmological simulations, such as the Millenium Project, which encode current understanding of the world model, of the clustering evolution of dark matter and of the physics of galaxy formation. By enabling inclusion of the most realistic models of galaxies, clusters and other structures, together with relevant analysis tools, VirtU will create a powerful resource to service the research needs of a wide community of theorists, observers and phenomenologists. VirtU will encompass a wide range of simulations, including, for example, simulations of star and planet formation. VirtU is an essential complement to AstroGrid and will form a key partner to the Virtual Observatory (VO) by providing capabilities to enable direct and rigorous comparisons of the best theoretical modelling with observational data. By enabling the link to be established between astronomical observations and astrophysical theory, VirtU will maximize the scientific returns from large observatories.

VirtU adds services and capabilities that are required to handle the unique requirements of the theory community. For instance a data model to handle theoretical 'particle' data, a compute grid to support mass scale generation of a zoo of galaxy models and 'grid−enabled' applications to manipulate these distributed data and information resources.

AstroGrid−2 interfaces with VirtU in a number of key areas. AstroGrid−2 exposes and enables access to observational data sets and resources, plus the applications required to manipulate them. AstroGrid−2 also provides the baseline technological components which VirtU aims to adapt to provide the fabric of the TVO (such as the community access model, MySpace, etc. VirtU will interact with AstroGrid−2 through the IVOA in areas of standards creation, to ensure interoperability of the data structures and services provided through the TVO.

9.4 UK Solar System Virtual Observatory

The UKSSVO is a UK consortium whose aim is to cover the needs for the solar system (SS) community to make use of the widely different datasets available. The SS community currently tends to work quite separately with very separate science and datasets ranging from in−situ, remote sensing,sampling of planetary geology etc . It is widely recognised that the exciting science will be at the overlap such as linking explosions on the Sun to responses in the Earth's and other planets' magnetospheres, exploring the relationship between physical processes such as magnetic reconnection on a wide range of scales. Hence the purpose of the proposal is to make use of the infrastructure developed by AstroGrid to enable the UK community to access the vast datasets available with ease. We will achieve this by providing input from the AstroGrid PM (Linde) and an AGLI (Harra) to the UKSSVO proposal.

There is a solar and STP component to AstroGrid, and hence the AstroGrid science requirements include relevant cases. This has meant that the infrastructure developed will be generic and can be used for a UKSSVO. This will also strengthen links between astronomy and solar system. It is natural for further development such as visualisation, infrastructure and feature recognition to have a generic component, and AstroGrid II and UKSSVO will collaborate on this. This will be described in the appropriate sections in the separate proposals. This aspect makes these projects unqiue in bring the astronomy and solar system communities together in a way that no hardware project, for example, can possibly achieve.

The aim of the UKSSVO is to develop a grid using infrastructure that will be available from AstroGrid in order to integrate all data from the UK solar system science community within a common framework. These exists a very diverse set of instruments, data products, and models, and the goal is to achieve the ability to share data and expertise across the boundaries of conventional scientific communities. The areas cover solar, heliospheric, interplanetary, magnetospheric, ionospheric and planetary. The initial challenge will be to obtain the same standards in all the datasets to allow these to be fed into the grid system, and to develop software to enable this. There will be specific tools developed such as visualisation tools. These will be researched alongside AstroGrid−2 to ensure that all generic software developed can be used to its full potential.

9.5 eSTAR: Grid Enabling Telescopes

The eSTAR project has developed prototype intelligent agents which control observing programmes, adapting them to the results of the observations thus far. They can only do this by interpreting the new data in the context of results from data mining. Thus, the eSTAR project have developed an expertise in using intelligent agents for data mining. Whilst data mining applications clearly lie within AstroGrid's remit, we felt that real time interaction with telescopes does not. Thus the eSTAR bid has been divided into two parts. Within this proposal we have applied for the resources to develop an intelligent agent toolkit which any astronomer could use for data mining. In a separate proposal (PI Iain Steele) we are applying for the resources to extend this toolkit to enable the agents to interact with telescopes, and to grid enable those telescopes.

9.6 Exploitation of ESO data in the UK

This proposal is submitted by RAL. It is a smaller e−science proposal than many others but will nevertheless provide significant benefits to the UK astronomical community by leveraging existing software and expertise.

Because UK astronomers are at present relatively unfamiliar with ESO instruments and the associated data processing techniques, there is a barrier to the UK's exploitation of data in the ESO archive. The aim of this proposal is to overcome this barrier by providing software which encapsulates knowledge of ESO instruments and the data processing steps required. The tools produced will use AstroGrid infrastructure where appropriate. Because RAL is also a member of the AG2 consortium, sharing personnel and expertise will ensure the software is fully compatible with AstroGrid.

9.3 VirtU: Theory Side Services

10. PROJECT STRUCTURE

Running such a complex and widely distributed project is extremely challenging. We have developed a firm but flexible management style for AstroGrid which will also be the basis for AstroGrid−2. This section describes the structure of the consortium and our management processes. Other elements of our project plan are found elsewhere in the proposal. Milestones and deliverables and required resources are specified in each workpackage section. Financial information, and our risk analysis, are detailed in appendices.

10.1 Project Constitution

The AstroGrid consortium will be expanded in this proposal to include new member institutions and collaborations with other projects. Additionally, this proposal will include some semi−autonomous R&D projects within its brief. All of this necessitates a new approach to the management of the project.

The ***consortium*** will include all institutes participating in AG2. It is understood that some members may not have staff deployed at their institute for all or part of the funding period (but the project will benefit by the involvement of their AGLI representative).

The ***AGLI*** (http://wiki.astrogrid.org/bin/view/Astrogrid/LeadInvestigators) will include one representative from each institute within the consortium, the PI for that institute. Meetings of the AGLI will be held at least monthly, either in person or via telecon.

The ***management team*** of Project Lead, Project Scientist and Project Manager will remain the same. These people will be in constant contact. The Project Lead will be elected from amongst the AGLI members as for AG1.

The ***TSP*** (http://wiki.astrogrid.org/bin/view/Astrogrid/TechnicalSupportPanel) was constituted in Phase B of AG1 to determine the priorities for each quarterly iteration of software development. This group will be reconstituted in AG2 as follows:

- Project Manager
- Project Scientist
- Technical Lead
- AVO/Euro−VO rep
- 1 x rep from main dev centres
    - Cambridge
    - Edinburgh
    - Leicester
    - RAL
- 1 x rep from major associated projects
    - SSVO
    - TVO
    - VEGA

We will constitute a group to oversee the major R&D projects, the AstroGrid Research Panel, ***ARP***, consisting:

- Project Lead
- Project Manager
- Project Scientist
- 1 x rep from R&D projects
    - ADEF
    - Intelligent Agents
    - Visualisation
    - AstroOntology
    - Scripting
    - Grid

The ***AGSAG*** (http://wiki.astrogrid.org/bin/view/Astrogrid/AGSAG) was constituted in Phase B of AG1 to provide scientific advice and direction to the project through the Project Scientist. This group will continue in its present form, though its membership may change to reflect the changed priorities of this new project. For instance, it may possibly include a science representative from the Euro−VO and NVO.

The project will continue to be overseen by the ***AGOC*** (http://wiki.astrogrid.org/bin/view/Astrogrid/OversightCommittee).

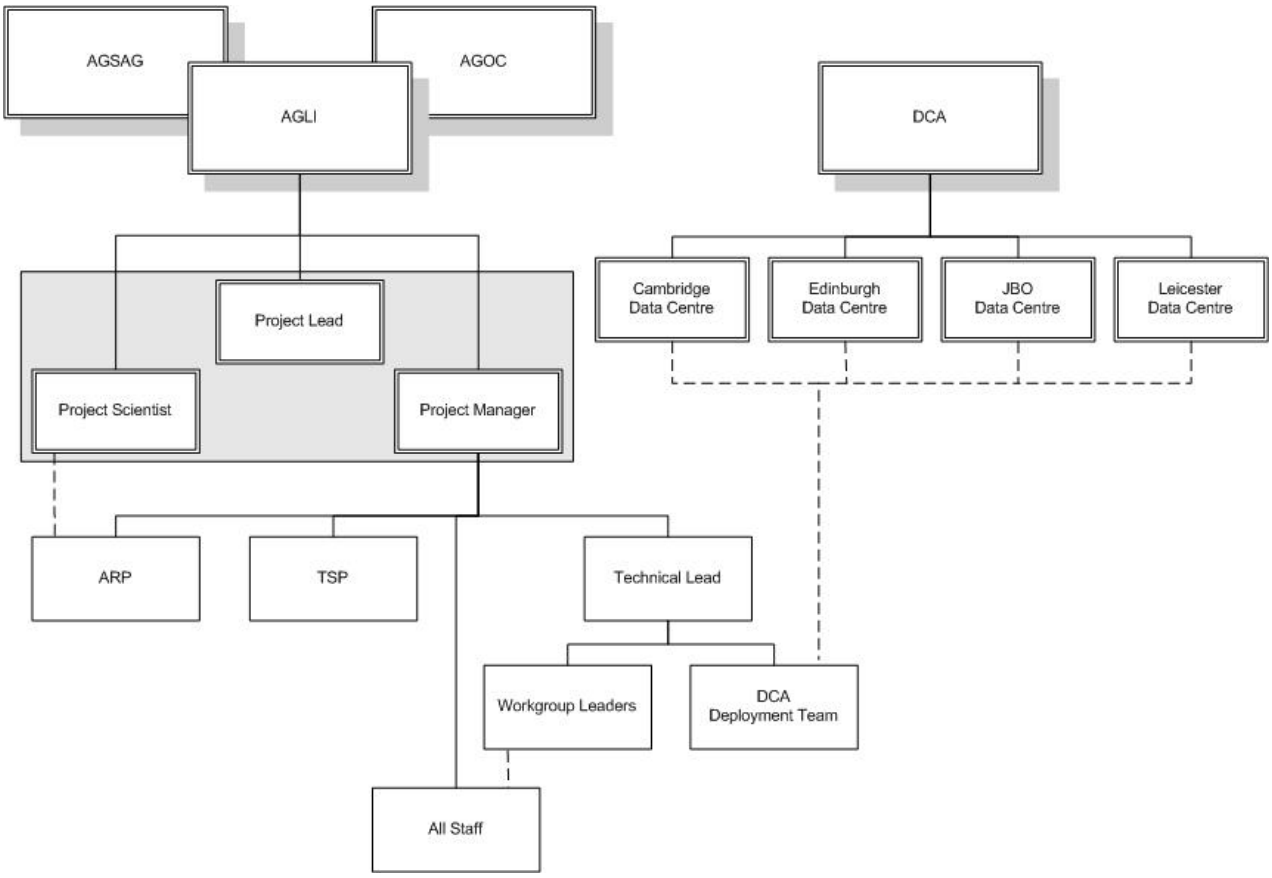
The ***DCA*** is a new initiative (see section 7.3) to roll out AstroGrid technology to key astronomical data centres in the UK and to ensure those data centres are properly staffed and maintained. AG2 will include a small roving team of developers who will assist the data centre personnel in deploying VO technologies and applications and act as a conduit for new, data centre specific infrastructure developments.

10.2 Project Management and Web presence

The conduct and control of AstroGrid−2 will continue in the same manner as in the original AstroGrid project, only taking account of the expanded role of the project (new partners, significant R&D sub−projects, inclusion of the Data Centre Alliance).

10.2.1 Project Management

The consortium section (10.4 'Consortium and Personnel') describes the new constitution of the AG2 consortium and its member institutes. The following chart shows the groups described in that section and their reporting relationships:



The AGLI is in overall control of the project. The Project Lead (elected from the AGLI) exercises that control on a day−to−day basis along with the Project Manager and Project Scientist. The AGOC is a PPARC appointed committee which reviews project progress and reports back to the GSC; the AGSAG is an advisory body set up to provide guidance to the project, primarily through the Project Scientist.

As the chart shows, all staff have a reporting line into the project manager but, in practice, this is exercised through various other individuals and groups. The ARP and TSP will determine the activities, on a quarterly basis, of the various R&D sub−projects and the infrastructure development teams. For day−to−day management, the Technical Lead and the workgroup leaders will control the development teams, while the R&D teams will be self−controlling. The DCA deployment team will be controlled jointly by the PM and representatives of the DCA.

10.2.2 Iterative and Incremental Development

We will continue the iterative and incremental approach adopted in AG1. ***Iterative*** refers to the practice of conducting the project as a series of iterations, mini−projects of three months length. ***Incremental*** refers to the idea that features are added to the end software incrementally, rather than trying to add every feature of a component in one big−bang.

For all ***infrastructure and tools development***, the content of an iteration will be determined by the TSP some four weeks beforehand. Workgroups will be created and leaders assigned to these. The leaders will be responsible for design, planning and estimates for their area and for managing the progress of the work, under the overall management of the Technical Lead.

All ***R&D*** projects will likewise be conducted on an iterative basis. Each project will present quarterly plans, including milestones and deliverables, to the ARP some four weeks beforehand. The ARP will approve or alter these plans and personnel assignments. Each project will also be expected to incrementally release specifications of enhancements to the VO infrastructure for consideration by the TSP.

10.2.3 Web Presence

One of the outstanding successes of the AG1 project has been its innovative use of web−based collaboration tools to make the conduct and progress of the project completely open. So successful has this been, particularly the use of the ***wiki***, that many other e−Science projects in the UK and Europe have followed our lead. These tools are:

- Portal: http://www.astrogrid.org
- News: http://news.astrogrid.org
- Forum: http://forum.astrogrid.org
- Wiki: http://wiki.astrogrid.org

One drawback of our present suite of tools is that each is completely separate, requiring the user to remember different sets of login information and different modes of creating and editing pages. We have started a project to replace the portal, news and forum aspects of the present site with a single integrated tool. This will be based upon Apache Cocoon technology (http://cocoon.apache.org/2.0/), the basis of our portal site (http://www.astrogrid.org/). We will propose to continue this effort into AG2, eventually extending the effort to add wiki capability.

We will maintain the current level of resourcing, 0.5 sy pa, to cover support and maintenance of the existing site and development of these new technologies.

10.2 Project Management and Web presence

APPENDICES

These appendices to the AstroGrid−2 proposal include those recommended by the Projects Peer Review Panel (PPRP):

Section−A : WORK PACKAGE DETAILS
Section−B : FINANCE DETAILS
Section−C : GANTT CHART
Section−E : RISK ANALYSIS TABLE
Section−F : PERSONNEL
Section−G : COLLABORATORS
Section−H : PUBLIC OUTREACH STATEMENT

together with information ancillary to the main proposal:

Section−I : DATA CENTRE ACTIVITIES
Section−J : OTHER ASSOCIATED PROJECTS
Section−K : ASTROGRID PUBLICATIONS
Section−L : REFERENCES

Section−A : WORK PACKAGE DETAILS

Management of the AstroGrid−2 project would follow the model used in the AstroGrid project, so finances and core development activities will be centrally managed, and all activities − both development and research types − will be subject to the quarterly review/enact/release cycle.

We can identify work packages as follows:

1. Management
    - ♦ Project Management
    - ♦ Science Requirements
    - ♦ Project Web and Collaboration Tools
2. Infrastructure Development
    - ♦ New Components
    - ♦ Maintenance, Refactoring & IVOA Cover
    - ♦ Science Tools
    - ♦ VO Personalisation Programme
    - ♦ Integration of R&D Prototypes
3. Research & Prototyping
    - ♦ AstroGrid Data Exploration Framework
    - ♦ Intelligent Agents
    - ♦ Visualisation
    - ♦ AstroOntology
    - ♦ Scripting Environment
    - ♦ Continuing Grid Research
4. Data Centre Alliance
    - ♦ VO Deployment
    - ♦ Training and Help Desk
    - ♦ Data Centre Staffing

Details of staff numbers (by FTE and SY) for the above are provided in the finance summary and in more detail in a separate spreadsheet supplied with this proposal.

Section−B : FINANCE SUMMARY

Detailed finance information is available in a separate spreadsheet. The following table summarises that information.

The PPRP format requires staff effort by institute, financial year and by work package. In the case of this proposal, only a few staff have been allocated to institutes: these people are named in the spreadsheet. The AG2LI will determine which of the current AstroGrid staff will be retained and where staff are required at other institutes when more detailed plans are drawn up following the allocation of funds from PPARC. However, in order to accurately model the likely salary costs in the attached spreadsheet, we have used figures derived from existing AstroGrid staff, including annual increments at 3%.

AstroGrid−2 Cost Breakdown			
	FTEs	***SYs***	***Total Cost***
01. AG2LI, Management, Admin & Web Dev	4.6	12.5	975,993
02. Infrastructure	11.0	33.8	2,211,385
03. Research & Prototyping	10.3	30.5	1,988,895
04. IVOA & Grid Involvement	3.0	9.3	618,821
05. Outreach Support	1.0	3.0	225,385
06. DCA: VO Support	4.0	5.0	293,623
07. DCA: Data Centres	8.0	30.0	1,766,358
Total Personnel	**41.9**	**124.1**	**8,080,460**
08. Travel			385,500
09. Personal Equipment			282,800
10. Project Capital Equipment			150,000
11. DCA Capital Equipment			300,000
12. General			50,000
Total Non−Personnel			**1,168,300**
Total			**9,248,760**

B.1 Line item explanation

1. AG2LI, Management, Admin & Web Dev
   includes 10% FTE for RAL LI plus Project Manager, Project Scientist, Technical Lead and Web Developer (50% FTE)
2. Infrastructure
   core development staff
3. Research & Prototyping
   research and development staff: new technologies for VO deployment
4. IVOA & Grid Involvement
   research and development staff: grid and IVOA standards involvement
5. Outreach Support
   developer to create specialist outreach software tools
6. DCA: VO Support
   staffing of deployment team, trainer and help desk
7. DCA: Data Centres
   staffing at (per data centre): manager (50% FTE), data manager (50% FTE), system admin
8. Travel & Equipment
   see below
9. General
   miscellaneous purchases

B.2 Travel Calculation

We will calculate travel using the following criteria (based on AstroGrid travel expenditure approved by Grid Steering Committee for Phase B):

Staff type	***UK Travel (pa)***	***Int'l travel (pa)***	***Number Staff***	***Total***

AG Management	£3.0K	£3.0K	4	£72.0K
IVOA contributors	£2.4K	£1.5K	5	£58.5K
Researchers	£1.2K	£1.5K	5	£40.5K
Senior Developers	£2.4K	£0.5K	5	£43.5K
Developers	£1.2K	£0K	12	£43.2K
Support Staff	£2.4K	£0K	4	£28.8K
DCA staff	£1.0K	£0K	12	£36.0K
AG2LI (excl PL)	£0.6K	£1.5K	10	£63.0K
TOTAL			57	£385.5K

AG Management includes Project Lead, Project Scientist, Project Manager, Technical Lead.
IVOA contributors includes personnel involved in IVOA working groups.
Researchers are personnel mainly employed in research activities.
Developers are personnel mainly employed in software development activities.
Support Staff are the DCA deployment team, Training staff and Helpdesk person.
DCA staff are institute−based personnel working on data centre activities.
AG2LI are the lead investigators for AstroGrid−2 project.

Personnel who fit into two or more groups are listed under the group with the greater travel allocation.

B.3 Personal Equipment Calculation

All personal equipment will be centrally funded from the budget held by the Project Manager (as with the current AstroGrid project).

For ***each person*** employed on the project we have budgeted for:

- £2.5K for personal equipment:
  (*project−lifetime cost*)
- £1.0K for additional equipment (peripherals, etc):
  (*project−lifetime cost*)
- £1.0K pa to institutes for equipment maintenance:
  (*annual cost, pro−rata on %FTE*)

Assuming the above number of personnel (excluding 11 AG2LI), this requires:

- (£2.5K + £1.0K) * 46 (= £161.0K) ***plus***
- £1.0K * 3 yrs * 40.6 (= £121.8K)

Total personal equipment budget is therefore ***£282.8K***.

B.4 Capital Equipment

The following major items of equipment will be required:

- for ADEF (section 6.1): three grid−enabled computational clusters with 10 Tb storage
  Cost: £100K
- for Visualisation (section 6.3): one commodity cluster with 2−3 Tb storage optimised for 3D image creation and streaming
  Cost: £50K
- for each of 4 data centres (see section : networking, hardware and storage equipment costed at £75K each
  Cost: £300K

Section−C : GANTT CHART

This section will be provided when the funding allocation from PPARC is known. Plans (to the half−year level) for each of the work packages are included in the main text.

Section−D : MILESTONES

This section will be provided as a composite whole and in more detail when the funding allocation from PPARC is known. Milestones for each of the work packages are included in the main text.

Section−E : RISK ANALYSIS TABLE

In this section we assess the likelihood and impact of major risks upon the AstroGrid−2 project and comment on the measures taken to avoid or offset those risks.

Risk	***Prob'y***	***Impact***	***Comment***
Technical			
Particular Grid technologies are not widely adopted	M	H	This is always a problem with leading edge technology projects We have instituted a detailed program of research into grid technologies, are closely aligned with the GGF so can track emerging trends and have close links with industrial partners (IBM, Oracle, Microsoft) so can assess emerging commercial trends
Functional			
Components do not integrate	L	VH	This is a serious problem in all IT projects: all efforts in the project are coordinated through review panels with a large number of common members so that problems such as this should be identified early; but the key tactic against this risk is that ***all*** development for the working VO infrastructure is carried out by the infrastructure team to ensure that components are well integrated; finally, the process of incremental development also mitigates against this risk as software is continuously being integrated and tested so problems are caught and dealt with early
Research in a field is a dead−end	VL	M	The work being undertaken in R&D (section 6) is far from *blue−skies*; most is a matter of applying techniques developed in other fields or extending current techniques to apply to new types or volumes of data. The impact identified is more on the personnel employed than on the overall usefulness of the AstroGrid−2 VO; the personnel concerned would be deployed in other areas.
Operational			
End system is non−performant	L	H	The AstroGrid VO is designed and developed as sets of interoperable components, each of which can be co−located or separate from other components so poor performance can sometimes be addressed by moving a component to a faster server; software is continuously integrated and tested so specific problems in performance can be identified and fixed early. What we will ***not*** do is spend months on performance tests as this would be futile in such a fast changing technology field
Social			
Lack of takeup in the community	M	H	Ensure that there is regular communication with the potential user community so that they are both aware of what will be delivered in the near future and they are aware of fora (AG science advisory group) for feeding in their needs.
Personnel			
Senior personnel are lost to project	L	M	PL: replace by other AGLI member PS: replace by senior scientist on project PM: replace by TLd and devolve some tasks to other senior members TLd: spread responsibilities among other Snr Devs
Research personnel are lost to project	L	H	spread knowledge of research work across institutes and ensure latest knowledge always on wiki
Development personnel are lost to project	M	L	do not concentrate knowledge of component in any one person or location
Difficulty in hiring staff	M	H	for researchers, we will begin advertising positions well in advance of the effort requirement; many of the development positions can be filled from existing AstroGrid staff the current downturn in the IT industry and the high−profile technologies that we are using mean that this is a very low risk for open developer positions applications for developer positions in AstroGrid have been encouraging (typically 50+)
Time & Money			
project is late	VL	H	with iterative development, project stops at given date

project is over budget	VL	H	as above: budget will be controlled so that overspending does not happen
costs of h/w or s/w go up	M	VL	project does not rely on outside software; hardware is a minor part of the budget
Activity Dependencies			
Lateness in one activity stops work in another field	L	H	The process of quarterly review/enact/release iterations will minimise the impact of such a possibility; we also benefit from having a wide range of activities under the AstroGrid−2 umbrella project − staff waiting on the completion of one activity can be usefully deployed on another activity

The assessment of probability and impact above are:

- ***VH***: very high
- ***H***: high
- ***M***: medium
- ***L***: low
- ***VL***: very low

The list of risk areas above are:

- ***Technical***: risks which arise from the technologies chosen, eg hardware, software, network
- ***Functional***: risks which arise from the components or architecture
- ***Operational***: risks that derive from the delivered system in operation, eg performance
- ***Social***: risks arising from the communities with which the project is concerned or involved
- ***Personnel***: risks arising from the personnel involved in the project, eg recruitment, loss of staff or skills
- ***Time & Money***: risk that we exceed budget or timescales
- ***Activity Dependencies***: risks arising from activity dependencies

Section−F : PERSONNEL

Apart from the Lead Investigators named in the next section, the following staff are named within the proposal:

Name	***Institute***	***Role***
Dr Nicholas Walton	Cambridge	Project Scientist
Tony Linde	Leicester	Project Manager
Keith Noddle	Leicester	Technical Lead
Dr Alasdair Allan	Exeter	Researcher
Elizabeth Auden (60%)	MSSL	Researcher
Dr Robert Mann	Edinburgh	Researcher
Dr Clive Page	Leicester	Researcher
Dr Anita Richards	JBO	Researcher
Dr Guy Rixon	Cambridge	Researcher
Tim Goodwin (50%)	Leicester	Web Developer

Section−G : COLLABORATORS

The AstroGrid−2 consortium represents the lead providers of astronomical data and information resources to the UK astronomical community.

Institute	***Lead Investigator***	***Notes***
School of Computer Science, Queens University of Belfast	Prof Fionn Murtagh	
Department of Physics, University of Bristol	Dr Malcolm Bremer	
Institute of Astronomy, University of Cambridge	Dr Richard McMahon	
Institute of Astronomy, University of Cambridge	Dr Nicholas Walton	AstroGrid Project Scientist
Institute for Astronomy, University of Edinburgh	Prof Andrew Lawrence	AstroGrid Project Leader
School of Physics, University of Exeter	Prof Tim Naylor	
School of Computing, University of Leeds	Prof Ken Brodlie	
Dept of Physics and Astronomy, University of Leicester	Dr Mike Watson	
Dept of Physics and Astronomy, University of Leicester	Tony Linde	AstroGrid Project Manager
Mullard Space Science Laboratory, University College London	Dr Louise Harra	
Jodrell Bank Observatory, University of Manchester	Dr Simon Garrington	
Institute of Cosmology & Gravitation, University of Portsmouth	Dr Robert Nichol	
Space Science and Technology Department, Rutherford Appleton Laboratory, CCLRC	Dr Peter Allan	

Section−H : PUBLIC OUTREACH STATEMENT

AstroGrid is a UK project which is part of a world−wide drive towards the concept of the ***Virtual Observatory***. The idea is that astronomical data from all over the world (pictures of the sky, catalogues of objects, spectra, radio maps, solar flare histories, and so on) should all speak the same language, and be transparently accessible and analysable using a uniform set of tools. Thus we will create a kind of *Digital Sky*. Today when you look at web pages with your web browser, it feels as if all those documents are sitting there on your PC. In just the same way, our ideal is that all the astronomical databases in the world should simply feel as if they are sitting on the scientists' workstation. Furthermore, by joining together computers in different cities in a ***Grid***, one can make them collaborate to perform elaborate calculations. For the individual scientist, it will feel like they have a *supercomputer on their desktop*.

The original AstroGrid project is scheduled to complete in December 2004. We are now planning a follow−on project, ingeniously called AstroGrid−2. This will extend and complete the original vision, but go further to develop new technologies in advanced ways to analyse data on−line and extract the scientific information from it, and in methods to robotically discover information spread across the world's databases.

AstroGrid is primarily aimed at professional researchers, but there are also exciting possibilities for the public, who can view the sky from the comfort of their homes, and for school students and teachers, who can plan and undertake real research projects with real data. We aim to develop software that will help school researchers.

AstroGrid−2 is a collaboration between eleven UK universities and laboratories.

Web page

http://www.astrogrid.org

Contacts

Project Leader : Andy Lawrence, University of Edinburgh, al@roe.ac.uk 0131−662−9736

Project Scientist : Nic Walton, University of Cambridge, naw@ast.cam.ac.uk, 01223−33−7503

Project Manager : Tony Linde, University of Leicester, ael@star.le.ac.uk, 0116−223−1292

Section−I : DATA CENTRE ACTIVITIES

I.1 Cambridge Data Centre Activities

The CASU Data Archive Centre is currently the single repository for nearly all of the raw optical and infrared data from the UK's ground−based telescopes. This includes:

- ***The archives of the ING***, which is a nearly complete record of all observations on those three telescopes dating back to 1984. Most of these observations are stored off−line on tape or CD.

- ***The UKIRT archive***. This was resurrected by JAC in 1997. An initiative to have these data served and supported in Cambridge was launched as a lack of support meant it would be impractical for the data to be served locally. The data are made available on tape and transfered to DVD at CASU. These are stored on−line in a DVD tower.
- ***The AAT archive***. We are currently mirroring the archive catalogues. The actual raw data are now being copied to DVD and will be shipped and installed at CASU by the end of summer 2003.
- The ***APM POSS1*** and ***UKST Sky catalogues***
- The ***INT Wide Field Survey***. The raw data are stored in the main ING archive. These are reduced and put on line as soon as they reach CASU. Currently the reduced data archive consists of 1Tb in images and 60 Gb in catalogues. This was recently used in an Astrogrid related exercise by implementing a SkyNode Web service to federate these data with SDSS, 2MASS and FIRST. This was a trans−atlantic federation and was singled out at ADASS−2002 as an excellent example of the way forward for the IVO.

Many commonly requested catalogues are also served from CASU. Some of these are USNO−B, Hipparcos, Tycho, CMT. In addition we have the ***UK's mirror copy of the CDS's VizieR service***, which contains roughly 3700 catalogues and is one of the most useful tools in all of astronomy today.

With the advent of affordable Tbyte online storage systems, it is now technically feasible to place all of the UK ground−based archives online, enabling automatic immediate access to all the data and the ability to correct faulty header information and thereby improving the internal consistency of the datI. A major project is currently underway to do just this for the ING data archive. We expect to have completed this task for the JKT and the WHT by the end of 2003, with the INT to follow soon in 2004.

Having all these data on−line is the first step towards the ultimate goal of on−the−fly calibration of data from any of the archives. The next step is to build upon exisiting data processing skills and further develop a unified strategy for automatic modular data processing pipelines, whether direct from archives or as part of dedicated survey systems. The ambitious goal is to optimally extract calibrated astronomical information in the primary data products. Any framework for doing this has to provide a means for incorporating accurate real−time astrometric, photometric and spectroscopic calibratation, in an automated way, if the data product is to be of general use. Included in this should be data quality control measures for assessing telescope, instrument and data pipeline performance. The deliverables would be quality−assured, calibrated images, spectra and catalogues delivered on−demand to the astronomical community over the internet.

As part of our data processing activities we have successfully built and run pipelines for many optical and IR imagers from around the world. Our current and most ambitious processing project is the ***pipeline for the UKIRT WFCAM imager***, which will be commissioned early in 2004. This pipeline will provide near real−time estimates of data quality at the summit as well as complete instrument signature removal and advanced processing capabilities for the data once they are shipped to CASU.

The synergy between the WFCAM and VISTA projects and the commonality of their respective requirements, strongly suggests that similar strategies should be implemented for the ***VISTA data flow system (VDFS)***. The processing modules written at CASU will be run at both the summit and at ESO to provide data quality control information. The fundamental and advanced pipelines will be run in Cambridge to provide the final data products. This is seen within the context of the longer term VEGA programme, a UK multi−institutional partnership, where processing techniques are developed in a systematic fashion to meet the increasing demands from missions such as Eddington and GAII.

Cambridge is one of the founding e−Science Centres, works closely with it and has representation on its management board. Cambridge also has frequent interaction with partners such as the Microsoft Cambridge Research Centre.

I.2 Edinburgh Data Centre Activities

The Edinburgh University Wide Field Astronomy Astronomy Unit (WFAU) concentrates on selected large scale surveys, and "added value" science archives. It is increasingly collaborating with local and international academic and industrial computer scientists on database curation, datamining algorithms, and grid infrastructure, leading towards a vision where an astronomical data centre is not just a repository, but a scientific analysis service centre. As well as development of advanced interfaces, implementation of algorithms, and design and tuning of experimental hardware solutions, this requires a very high standard of robustness, calibration, documentation, and continuing curation.

The main data products which the WFAU is currently hosting or developing are :

- The ***UK Schmidt Plate Archive*** Over 17,000 plates of lasting value from the UK Schmidt, Palomar, and other sources.

- ***The SuperCOSMOS Science Archive (SSA)*** A 4TB all−sky multi−band multi−epoch sky survey based on scans of plates from the UK Schmidt and Palomar. It also includes the ***SuperCOSMOS H−alpha Survey*** of the Galactic Plane. Pixels and catalogues have been on−line for some time, but we are about to release a much more advanced queryable interface based around MS SQL Server.

- ***The Sloan Digital Sky Survey*** We host a mirror of the SDSS−EDR based on *Objectivity*, an object oriented database. A mirror of DR1 (using SQL Server) will also be established.

- ***The 6dF Galaxy Redshift Survey*** A queryable spectroscopic database linked to 2MASS and SuperCOSMOS imaging datI. It runs both a push−button interface and a write−your−own−SQL option.

- ***The WFCAM Science Archive (WSA)*** This is currently under development, in collaboration with CASU who run the pipeline, while we ingest the calibrated data and serve it to external users. The WSA is the prime UK astronomy database until VISTA arrives, not just because of its scientific interest in producing the first IR equivalent of the optical sky surveys, but also because of its size (20TB a year), and its timing − the first archive specifically designed to be VO−ready. The WSA will be release in 3 versions with increasingly ambitious interfaces.

- ***The VISTA Science Archive (VSA)*** The WSA is seen as the testbed for VISTA, which will be a prime ESO facility, and deliver IR survey data at a rate of 100TB/year. We need to be ready to serve VISTA data to users in 2006. The VISTA Data Flow System (VDFS) is now part of a larger whole, a multi−institution plan to work coherently towards VISTA, GAIA, and Eddington. (VEGA).

Further new databases are under consideration, but will always be linked to local scientific interests − for example the idea of an equivalent to UKIDSS for SCUBA2, or data from JWST−MIRI, and OWL.

The Edinburgh WFAU works closely with the National e−Science Centre (NeSC), which is an Edinburgh−Glasgow collaboration, and related activities such as eDIKT.

I.3 Jodrell Bank Data Centre Activities

Our main projects are

- Providing MERLIN data in a variety of formats adapted to the needs of the user, from partly−processed data for the radio expert, to a ready made image for rapid multi−wavelength comparison;
- Helping to establish standards and common procedures for easy VO access to any interferometry data;
- Laying the foundations for the VO−compatibility of e−MERLIN and ALMA data (to be continued via a separate bid).

The MERLIN archive catalogue lists all public domain data taken since 1992. This can be searched via a web browser (established in 2000). Since then, we have been building up an archive of processed data and images of the target objects. We are developing tools to extract a range of products from the calibrated visibility data, such as images at a range of angular scales and positions, spectra, radio `light' curves and even rotation measure images or visibility amplitudes for gravitational lens modelling. The user does not need any specialised interferometry knowledge.

At present continuum images representing 6 years of observations can be accessed on−line, either via the MERLIN web site or via the Vizier and Aladin catalogue and image browsers. The latter enables radio images to be compared directly with other images at a comparable resolution (e.g. from HST or CHANDRA), as used for the AVO First Light demonstration.

Our next steps are to publish the MERLIN archive in AstroGrid as a web service and extend the availability of products extracted on demand. Once this is achieved we can also host other data sets, either to exploit interferometry data processing routines (as we have already done for VLA and other data to combine with MERLIN projects), or to test and impliment the optimisation of AstroGrid queries using distributed resources.

We will also continue to populate the archive and take advantage of AstroGrid facilities such as authentication procedures to allow PIs to access proprietary data and using MySpace to cache on−demand images which take more than a few minutes to process. A spectral data cube or an ultra−sensitive observation can take hours or even days to process although we will reduce this by parallelization of data reduction.

This challenge will increase as first e−MERLIN and then ALMA come on−line. Our involvement in AstroGrid as well as in these projects means that VO compatibility is being designed into their archiving processes right from the start. We are already in detailed discussion with representatives of other European radio synthesis arrays (from EVN/JIVE and Westerbork) to facilitate access to their archives, as well as participating in the world−wide interferometry interoperability discussions lead by ATNF under the auspices of the IVOI. We have also been developing interferometry metadata since 2001 and we will continue to expand and refine this in response to VO and end−user requirements.

In addition, groups based at JBO are major partners in several other international projects involving the management of large data sets. Staff working on the projects listed below are already in discussion with AstroGrid and other IVOA partners to develop VO access to their data:

- The CERES gravitational lens database
- The Parkes Multibeam Pulsar Survey
- Planck and other CMB−related data including foregrounds
- The HIPASS and HIJASS all−sky HI surveys

This will involve handling many types of interferometry and single dish data including time series and spectral cubes. UMIST, which is in the process of merging with the University of Manchester, is responsible for a definative astrochemical database. Integrating the use of such databases into VOs, to provide spectral templates and test models, is one of the challenges which must be met to exploit molecular line data such as ALMA observations.

I.4 Leicester Data Centre Activities

An astronomical data archive centre was first established at the University of Leicester in September 1992. Currently Leicester provides, via the LEDAS service, an on−line astronomical database service and access to archive data from high energy astrophysics missions. LEDAS is sole UK provider of high energy archives, in particular LEDAS provides the primary means of access for the UK astronomical community to:

- the ***ROSAT Public Data Archive***: pointed and survey observations;
- the ***ASCA Public Data Archive***;
- the ***Ginga Products Archive***;
- the ***Chandra Science Archive***.

Each of these datasets includes observing logs, X−ray source catalogues, associated data products such as X−ray images, spectra, light curves (details vary from project to project) as well as the `raw' data files from each observation. In addition to these datasets, LEDAS also holds the archival data from earlier X−ray missions such as Einstein & EXOSAT.

LEDAS has recently added the XMM−Newton `1XMM' Catalogue and its associated data products to its collection and will incorporate further XMM−Newton catalogue releases as they become available.

In addition LEDAS also provides access to:

- a large collection of other astronomical catalogues from ground−based (primarily optical/IR/radio) observations and other space missions;
- a local interface to the Vizier service at CDS;
- tools for accessing the DSS;
- advanced interfaces to some of the largest existing catalogues (USNO A2, 2MASS and most recently the first UK access to USNO B1.0).

Currently archival data storage amounts to ~1TB of catalogue and archive data on magnetic disk, currently increasing at ~400 GB/yr. Usage of LEDAS amounts to 40,000 web page views and >1 GB downloaded per month

From its inception LEDAS has provided a full function database query and visualisation environment. In the decade or more that LEDAS has existed it has been at the forefront of developments in delivering on−line data to the community with powerful, integrated interfaces and advanced search and manipulation facilities, pioneering the approach that we now call the Virtual Observatory. This track record provides Leicester with extensive expertise on which to build in the AstroGrid−2 erI.

LEDAS has strong collaborative links with other data centres, notably with the HEASARC (NASA's High Energy Astrophysics Science Archive Research Center located at Goddard Space Flight Center, Washington DC) and CXC (the Chandra X−ray Observatory Center at the Harvard−Smithsonian Center for Astrophysics in Cambridge, Massachusetts).

In the near future LEDAS plans to develop in several areas:

- ***WASP***: Leicester will host the science archive from the WASP optical survey project. WASP will survey a large fraction of the sky nightly with the aim of discovering of extra−solar planets using the transit method as well as optical transients and near−earth objects. Significant developments in the area of time−domain data−mining problems are anticipated in the context of WASP.
- ***The UK SWIFT Data Centre***: Leicester will be one of the 3 international data centres for NASA's SWIFT satellite due for launch in December 2003. SWIFT will focus on the study of gamma−ray bursts and their afterglows with a complement of X−ray and UV/optical instrumentation. LEDAS will host the UK SWIFT data archive.
- ***The XMM−Newton XID Programme Public Database***: The XMM−Newton XID Programme is providing optical/IR identification data for large samples of XMM−Newton X−ray sources (e.g. from the INT WFC). LEDAS will host this key dataset for which a strong VO demand is expected.

A successful bid by Leicester resulted in its selection as one of the 7 new e−Science Centres of Excellence. Leicester's Centre of Excellence will focus on e−science problems in the astrophysics area and has strong links to existing local activities such as LEDAS, current Leicester AstroGrid activities and the UK Astrophysical Fluids Facility (UKAFF) project.

I.4 Leicester Data Centre Activities

Section−J : OTHER ASSOCIATED PROJECTS

J.1 UK e−Science Programme

J.1.1 NeSC & UK e−Science

The National e−Science Centre (NeSC) plays a key role in the development and progress of e−science in the UK. AstroGrid has enjoyed a very fruitful relationship with NeSC to date: AstroGrid members have been amongst the most regular and most active attendees at NeSC workshops; AG2 members lead all the first three e−science Special Interest Groups set up under the aegis of NeSC to date; AstroGrid has proposed, and greatly benefitted from, several training courses run at eSI in Edinburgh; and NeSC has funded an AstroGrid member, Bob Mann, in a half−time "Research Leader" position for one year, principally to liaise between NeSC and the VO community. A continuation of this close relationship is expected between NeSC and AG2, made stronger by the direct involvement of the NeSC−based eDIKT project in AG2 work, as described below. AG2 will also benefit from access to BlueDwarf, a 16 CPU, 128GB RAM SMP machine located at NeSC, for experimental data integration and data exploration work.

Other AstroGrid involvement in the UK e−science programme includes:

- Leicester eScience Cente of Excellence: specialises in astronomy − Watson is PI, Linde co−I
- Bristol eScience Cente of Excellence: specialises in e−Digital media − Bremer is a co−I
- Cambridge Regional eScience Centre: Walton is on its Board of Management.

Members of AstroGrid are also involved in a number of e−Science research programmes supervising students through the PPARC e−Science studentship scheme.

J.1.2 OGSA−DAI/DAIT

The AstroGrid consortium was amongst the most active contributors to the *Data Requirements for the Grid* scoping study which led to the setting up of the OGSA−DAI project to develop data access and integration services within the framework of the Open Grid Services Architecture. Since then, AstroGrid has been one of the two "early adopter" projects for OGSA−DAI, given access to early releases of software and providing feedback on them to the design team. This relationship has been valuable to both sides, and it is hoped that similar contact can be maintained between AG2 and DAIT, the successor to the OGSA−DAI project. OGSA−DAI is recognised as a considerable success internationally, and as its services become established as standards within the Grid community, they will become more important for the VO. DAIT will extend the functionality of the OGSA−DAI services, as well as their robustness, so it is likely that AG2 will have even more to gain from collaboration with DAIT than was the case for AG1 and OGSA−DAI.

J.1.3 eDIKT

The eDIKT project is a SHEFC−funded e−science initiative, based at NeSC in Edinburgh, which is intended to produce well−engineered solutions to e−science problems, using cutting−edge computer science research techniques. eDIKT's work is split into a generic Grid computing strand (currently producing a second Reference Implementation of the core OGSA−DAI services, as required by the GGF standardisation procedure) and a series of testbeds, driven by application scientists in Edinburgh and Glasgow. An active astronomy testbed is currently investigating the use of BinX, an XML schema for describing binary data, in the VO context: initially this is simply performing conversion between VOTable and FITS files, but further uses are planned. eDIKT is expecially interested in large−scale data management and integration projects, and is committing 1 FTE of software development effort to the AstroGrid Data Exploration Framework (ADEF), as described in Section 6.1.

J.2 European and International Initiatives

J.2.1 Euro−VO

AstroGrid is a partner in the Euro−VO project, which grows out of the AVO (Astrophysical Virtual Observatory project. AVO was funded by the EU through the FP5 RTD programme as a Phase A R&D study leading towards a more ambitious Phase B proposal to establish a working Virtual Observatory for Europe; Euro−VO is that Phase B proposal. AstroGrid has participated in AVO, gaining 3 FTEs of funding, and has carefully aligned its work, so that approximately 6 FTEs of effort can count both to AstroGrid and AVO simultaneously. For Euro−VO, this alignment of work is even more crucial.

Euro−VO plans the creation of three entities. (1) The VO Facility Centre (VOFC) will provide a co−ordinating point, registry maintenance, and an end−user support system, including continuing science requirements development. This will be a persisting entity based at ESO. (2) The Data Centre Alliance (DCA) focuses on backbone support to European Data Centres, especially in the areas of VO technology uptake, and Grid−node establishment. There are eight formal partners, but these are expected to act as *national nodes* linking to a wider list of data centres. (3) The VO Technology Centre (VOTC), a distributed organisation aimed at completion of the VO infrastructure, and development of new technologies. AstroGrid has a recognised role in leading the VOTC programme. This is organised into five themes : core VO completion; science user tool development; automated resource discovery; high volume datamining on the Grid; and visualisation techniques. It is ***not*** an accident that these technology themes, along with the DCA strand, bear a marked resemblance to the structure of the AG2 proposal ! Precisely how the combined VOTC/AG2 programme is carried out and managed will be finalised once the outcomes of PPARC and EU funding bids become clearer.

Because of the Byzantine intricacies of the FP6 programme, the above Euro−VO structure is being requested through three separate proposals in a kind of matrix. Two of these have been submitted and are available at http://wiki.astrogrid.org/bin/view/VO/EuroVOPlans. The third (for which A.Lawrence will act as co−ordinator) is expected to meet a call with October deadline. If these proposals are 100% succesful, the UK will gain somewhere in the range 6−9 FTEs for four years starting in late 2004. The proposals contain sizeable *co−ordinator reserves* which will be allocated to partners after one year. This is partly to allow the possible addition of further partners, but also to allow the allocation of these posts against *matched national funding* − in other words the more PPARC positions we can deploy, the more EU positions we will get.

J.2.2 iAstro

In 2001, a European COST ("Cooperation in Science and Technology") project was started, in the telecoms and information technologies area, with the title "iAstro − Computational and Information Infrastructure in the Astronomical DataGrid". This 4−year project organises workshops and research collaborations. Workshops have been held in Edinburgh, Strasbourg, and Granada, on themes which include: XML metadata standards for image databases, multispectral imaging, peer to peer prototypes, the curvelet transform, Bayesian methods in image restoration, independent component analysis, and other topics. Upcoming events are planned for Nice (Oct. 2003) and Capri (March 2003). iAstro has been very active in supporting FP6 submissions, and it has submitted a Marie Curie Research Training Network initiative. AstroGrid members involved in the Management Committee of iAstro are QUB (lead) and RAL. Further details are available at http://www.iAstro.org

J.2.3 Global Grid Forum

The Global Grid Forum (GGF) is the body defining standards within the international Grid community and therefore sets the framework within which much of AstroGrid's work taskes place. AstroGrid consortium members have attended several of the recent GGF meetings, and participate in GGF working groups. An AstroGrid member, Guy Rixon, is one of the chairs of the recently−established Data Format Description Language (DFDL) working group, which is defining an XML language for describing the structure of binary and character encoded files and data streams so that their format, structure, and metadata can be exposed: this work is an extension of the development on BinX, which itself has the requirements of the VO at its core, as described above.

J.2.4 International Virtual Observatory Alliance

The International Virtual Observatory Alliance (IVOA) grew out of occasional telecons in 2001 between the three then−funded VO projects − AstroGrid, AVO, and US−NVO. By summer 2001, several other projects worldwide were beginning, and a semi−formal alliance was developed, including an executive structure, and a *roadmap* written by Quinn, Hanisch and Lawrence. Since then the IVOA and its usefulness have both grown rapidly, with a web site at http://www.ivoa.net, working groups in technical areas, email discussion forums, and physical meetings several times a year, some being policy and planning meetings, and some being technical standard *interoperability* workshops. There are now 12 projects within the IVOA, and the current executive is developing new policy guidelines on how we go about our business. The IVOA is the main route by which international interoperability is maintained. It does this partly by being a discussion forum for exchange of both ideas and experience, but also increasingly as a formal body for standards proposal, review, and approval, equivalent to the the W3C. AstroGrid is strongly active within the IVOA and will maintain this level of activity. Nic Walton is the secretary of the IVOA, and in 2005 Andy Lawrence will take over the rotating chairmanship.

Section−K : ASTROGRID PUBLICATIONS

K.1 Publications

Garrington, S. T.; 2002, Proceedings of the XXV11th URSI General Assembly, 2002, URSI, 'The Future of Long Baseline Interferometry'

Genova, F.; Benvenuti, P.; De Young, D. S.; Hanisch, R. J.; Lawrence, A.; Linde, T.; Quinn, P. J.; Szalay, A. S.; Walton, N. A.; Williams, R. D., 2002, American Astronomical Society Meeting 200, #87.03, 'International Collaboration for the Virtual Observatory'

Gilmore, G., Walton, N. A.,, 2003, Frontiers, 16, 10, 'The Virtual Observatory'

Lawrence, A. and the Astrogrid consortium, 2002, Virtual Observatories. Edited by Szalay, Alexander S. Proceedings of the SPIE, Volume 4846, pp. 6−12, 'The AstroGrid Project : Powering the Virtual Observatory'

Lawrence, A., 2001, Virtual Observatories of the Future. Edited by Brunner, R.J., Djorgovski, S.G., and Szalay, A.S. ASP Conference Proceedings of the SPIE, Volume 225, pp. 114−116, 'Astro−IT challenges and big UK survey programmes : SuperCOSMOS, UKIRT WFCAM, and VISTA'

Mann, R. G. et al, The Astrogrid Consortium, 2002, Astronomical Data Analysis Software and Systems XI, ASP Conference Proceedings, Vol. 281. Edited by David A. Bohlender, Daniel Durand, and Thomas H. Handley, 'AstroGrid: the UK's Virtual Observatory Initiative'

Quinn, P., Benvenuti, P., Diamond, P.J., Genova, F., Lawrence, A., Mellier, Y., 2002, Virtual Observatories. Edited by Szalay, Alexander S. Proceedings of the SPIE, Volume 4846, pp. 1−5, 'Astrophysical Virtual Observatory (AVO) : a progress report'

Richards, A. M. S.; Garrington, S. T., 2003, Towards an International Virtual Observatory, Garching, 2002, Springer−Verlag Quinn, P. ed., 'Star−Forming Regions at High Resolution: Interferometry for Virtual Observatories'

Richards, A. M. S.; Garrington, S. T., Reynolds, C.; Allen, M. G.; 2003, The Scientific Promise of the SKA', SKA, Kramer, M.; Rawlings, S.; eds, 'Virtual Observing for the SKA'

Rixon, G. T.; Walton, N. A., 2002, Virtual Observatories. Edited by Szalay, Alexander S. Proceedings of the SPIE, Volume 4846, pp. 115−123, 'Identified usage of the virtual observatory: beyond the WWW'

Walton, N. A., 2002, Astronomy & Geophysics, Volume 43, Issue 1, p. 30, 'AstroGrid: Powering the virtual universe'

Walton, N. A.; Lawrence, A.; Linde, A. E., 2003, Astronomical Data Analysis Software and Systems XII ASP Conference Series, Vol. 295, 2003 H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds., p.25, 'Scoping the UK's Virtual Observatory: AstroGrid's Key Science Drivers'

K.2 Talks and Presentations

A complete list of presentations made by the AstroGrid team in the period 10/2001 to 05/2003 is given at http://wiki.astrogrid.org/bin/view/Astrogrid/ConferencePapers

K.3 Seminars and Talks

A complete list of seminars given by the AstroGrid team in the period 10/2001 to 05/2003 is given at http://wiki.astrogrid.org/bin/view/Astrogrid/SeminarsAndTalks

Section−L : REFERENCES

Allan. A, Naylor T., Steele I., Carter D., Etherton J., Mottram C., 2003, 'eSTAR: Building an Observational GRID', In, Proc ADASS XII, Payne H.E., Jedrzejewski R.I., Hook R.N. (eds), pp. 13, ASP Conf. Series 295, Astronomical Society of the Pacific, San. Fran.

(AstroVirtel): http://www.stecf.org/astrovirtel/ and accepted proposals at
http://archive.eso.org/wdb/wdb/vo/avt_prop/query

(AVO) The Astrophysical Virtual Observatory − a three year EC funded programme charged with mapping out the structure of a facility class virtual observatory for Europe. See http://www.eso.org/avo

Cavanagh,B., et al 'ORAC−DR: One Pipeline for Multiple Telescopes', In Astronomical Data Analysis Software and Systems XII ASP Conference Series, Vol. 295, 2003 H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds., p.237.

D. De Roure, N. Jennings, and N. Shadbolt. "Research Agenda for the Semantic Grid: A Future e−Science Infrastructure".
Technical report UKeS−2002−02, UK e−Science Technical Report Series,
National e−Science Centre, Edinburgh, UK. December 2001.

(EGSO) European Grid of Solar Observatories − an EU IST funded programme. See
http://www.mssl.ucl.ac.uk/grid/egso/

(Euro−VO) The current working title for the European Virtual Observatory initiative (shortly at http://www.euro−vo.org). AstroGrid is providing vital scientific and technical input into the development of this future programme.

The Great Observatories Origins Deep Survey (GOODS) is a public, multiwavelength survey that will cover two 150 arcmin2 fields. These fields are centered around the HDF−N (Hubble Deep Field North) and the CDF−S (Chandra Deep Field South): see http://www.eso.org/goods.

Gruber, T.R. (1993). "A Translation Approach to Portable Ontology Specification."
Knowledge Acquisition ***5***: 199−220.

The International Virtual Observatory Alliance. It's home page is at http://www.ivoa.net. The IVOA Mission and Roadmap is located at
http://wiki.astrogrid.org/bin/view/IVOA/RoadMap[http://wiki.astrogrid.org/bin/view/IVOA/RoadMap

Natalya Fridman Noy and Deborah L. McGuinness. "Ontology Development 101: A Guide to Creating Your First Ontology".
Stanford Knowledge Systems Laboratory Technical Report KSL−01−05
and Stanford Medical Informatics Technical Report SMI−2001−0880, March 2001.
http://www.ksl.stanford.edu/people/dlm/papers/ontology−tutorial−noy−mcguinness−abstract.html

(NAM2002) National Astronomy Meeting 2002 in Bristol, see http://www.star.bris.ac.uk/nam

Naylor T., Steele I., Carter D., Allan A., Etherton J., Mottram C., 2003, 'eSTAR: Telescopes and Databases as a Single Information Grid', In, Towards an International Virtual Observatory, Proceedings of the ESO Workshop held in Garching, Germany, 10−14 June 2002, Springer−Verlag, in press.

(NVO) The US National Virtual Observatory project: http://www.us−vo.org

Page L., & Brin S., 1998, 'The Anatomy of a Large−Scale Hypertextual Search Engine', In, Proceedings of the 7th International World Wide Web Conference, Elsevier Science.

(SDT) National Virtual Observatory Science Definition Team (SDT) − final report at http://nvosdt.org/sdt−final.pdf

(SpaceGRID) An ESA funded programme to investigate possible uses of Grid technology in supporting the scientific and technical functions of ESA: http://spacegrid.esa.int

SuperCOSMOS Sky Survey: http://www−wfau.roe.ac.uk/sss/

(UKIDSS) UKIRT Infrared Deep Sky Survey

Enabling Outreach with the NVO at http://bill.cacr.caltech.edu/cfdocs/usvo−pubs/files/NVO_EPO_Recs.pdf

(WFS) The Isaac Newton Group's Wide Field Survey programme: http://www.ast.cam.ac.uk/~wfcsur/index.php

(XMM−SSC) XMM−Newton Survey Science Centre: http://xmmssc−www.star.le.ac.uk/