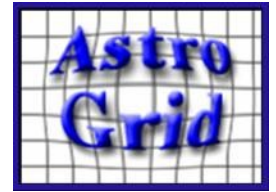


# AstroGrid-2 Project

CYCLE-5 PLAN : final version  
2007-03-09



## AstroGrid Management Team

Andy Lawrence	<i>AstroGrid Project Leader, University of Edinburgh</i>
Tony Linde	<i>AstroGrid Programme Manager, University of Leicester</i>
Keith Noddle	<i>AstroGrid Project Manager, University of Leicester</i>
Nic Walton	<i>AstroGrid Project Scientist, University of Cambridge</i>
Mike Watson	<i>AGLI representative, University of Leicester</i>
Richard McMahan	<i>AGLI representative, University of Cambridge</i>

---

## (1) Introductory Notes

AstroGrid-2 operates on a six month planning cycle, operating on calendar semesters. There is an overall [Lifetime Roadmap](#). For each cycle however we produce a top-level "Cycle Plan" (this document) which is fixed for that cycle. This is used by the AG management team to track progress at fortnightly meetings, by the AGLI at quarterly meetings, and by PPARC's AstroGrid Oversight Committee (AGOC) to review progress at six monthly meetings. Within a cycle, development processes are much shorter and very agile, but the top level goals will not be revised during the cycle. (Any changes required by circumstances will simply be reported to the AGLI and AGOC). In parallel with developing this *Cycle Plan* the Project Manager works with the project team to develop a *Detailed Project Plan* to implement the Cycle Plan goals. The Management Team are responsible for producing the plan, following planning meetings involving project staff, and subject to AGLI approval.

## (2) Goals for this Cycle.

The over-arching aim of Cycles 5 and 6 is to make AstroGrid software and working systems ready for the UK VO operations service planned to start in 2008. This is a plausible aim because at this stage we have a complete core software suite, a good user interface, and increasing takeup by data centres, by other VO projects, and by end users. However the system is not yet robust or user friendly enough to the standard we intend. We therefore need to be disciplined in these final two cycles, minimising new functionality, and technological changes. (Technological changes cannot be completely avoided because of continuing IVOA developments). The prime goal is therefore :

- to make the existing AstroGrid system **robust, working, and useful**

The top-level **goals** for Cycle-4 were (i) to improve software performance, robustness and reliability (ii) to establish a wider range of data access (iii) to establish a wider range of compatible tools. Goals (ii) and (iii) were successfully achieved; progress was made on goal (i), but it was not achieved to the desired standard. This goal is crucial to the success of the project, and therefore *system performance* is the highest priority for this cycle. Checking through the **deliverables** from the last cycle, most were successfully delivered, but *user documentation* remains incomplete. This is currently a significant barrier to adoption and useability, and is then the next highest priority. Very good progress was made in Cycle-4 on data access and third party user tools, but feedback at workshops has produced a clear list of existing applications that users wish to see wrapped within

the AstroGrid system. The third priority is therefore to *integrate more applications*. Finally, comparing progress against the eight top-level goals of the original AstroGrid proposal, we find that we have scored very well, but agree that we need to place most weight on one goal : "to produce a working VO system of real scientific use to astronomers". Our frank assessment of this is as follows. Astronomers at workshops and other events are impressed by what they see, and after a little exposure are convinced that this is indeed the way work will be done from now on. We have several hundred registered users, and astronomers using AstroGrid tools at workshops have achieved real scientific results. It cannot yet be said though that AstroGrid is the daily tool of choice. To a large extent this is because of the issues already described above : *system performance, documentation, and services offered*. However it is also likely that, now that the components work and we have seen how users interact with them, that a *re-think of the interface structure* is needed, to fit it as closely as possible to what astronomers want to do. We do not intend any radical overhaul this cycle, but will undertake a critical review, with a view to a refreshing the UI in Cycle-6 before final release.

These then are the Cycle-5 **goals in priority order** :

- to improve software performance, robustness and reliability
- to complete the user documentation
- to integrate more science applications
- to undertake a critical review of fitness for purpose of the UI

### (3) Deliverables list for Cycle-5

Specific deliverables are summarised here. Brief explanations are appended below. VOTECH deliverables are included - but note that these are not completely under the control of AstroGrid.

C5-01	Data Exploration Study report (VOTECH)		2007 Feb 15
C5-02	Infrastructure Study report (VOTECH)		2007 Feb 15
C5-03	Science Framework Document (VOTECH)		2007 Feb 15
C5-04	Reduced Science Workshop programme	Cardiff	2007 Jan 17
		Belfast	2007 Mar 29
C5-05	Second baseline s/w release (VOTECH)		2007 Mar 30
C5-06	Review/refactor Integration process		2007 Apr 01
C5-07	Deployed system release 2007.1		2007 Apr 01
C5-08	Set VOTECH Stage-5 goals		2007 Apr 04
C5-09	Review Tools Call No. 1		2007 Apr 15
C5-10	Demo at NAM		2007 Apr 16
C5-11	Produce compt. config/administration mechanisms		2007 May 01
C5-12	Complete documentation for science users		2007 May 01
C5-13	Conduct Fitness review		2007 May 01
C5-14	Complete documentation for technical users		2007 May 15
C5-15	Issue Tools Call No. 2		2007 May 15
C5-16	Meet IVOA commitments		2007 May 18
C5-17	Improve s/w performance/robustness		2007 Jun 01
C5-18	Complete Taverna prototype		2007 Jun 01
C5-19	Integrate applications "wish list"		2007 Jun 15
C5-20	Convert key workflows to Python scripts		2007 Jun 15
C5-21	Resource Browser		2007 Jun 15
C5-22	STILTS wrapping		2007 Jun 22
C5-23	Other component developments		various dates
C5-24	Develop key external collaborations		various dates

#### (4) Other Milestones for Cycle-5

Feb 01-02	DCA Board
Feb 05	AGSAG Feb 5th
Feb 07	AstroGrid Oversight Committee
Feb 08	Caltech/SDSC collab visit Feb 2007
Mar 06	AGDAG-4
Mar 12-16	VOTECH DS planning meeting
Apr 04	VOTECH Board
Apr 16-20	NAM07 UCLAN
Apr 17	AGLI quarterly (at NAM : provisional)
May 07-11	OGF20 Manchester
May 14-18	IVOA interop, Beijing
Mar 21-23	Euro-VO workshop ESAC
May 22	FP7 deadline
Jun xx	AstroGrid Oversight Committee (TBD)
Jun 11-15	Cycle-6 planning meetings
Jun 21-22	Full consortium meeting Edinburgh Jun
Jun 22	AGLI quarterly (at ConsMeet : provisional)
Jun 25-29	DCA workshop : "How to publish in the VO"

Formal AGLI meetings now occur quarterly rather than monthly; however Project Management Team meetings, including three AGLI members, occur fortnightly, and times are advertised so that other AGLI members may join in if they wish.

#### (5) Cross check with Project Lifetime Roadmap

Comparing with the original long term roadmap, we are fairly much on target. The following planned milestones deserve some comment :

2006 Jun Single Sign-on for VO services : AstroGrid implementation is ready; awaiting IVOA standards process.

2006 Oct Client side scripting for workflow : essentially achieved through AR Python library. Now need to achieve for other languages.

2006 Dec Federated access : VOspace. Making slow but steady progress on IVOA front.

2006 Dec Secure authorised access to VO services

2007 Jun Full access control. As above, awaits progress at IVOA.

#### (6) Brief explanations of deliverables.

**C5-01 (Data Exploration Study report)**. This is a VOTECH deliverable that was due last Cycle. It is part done and will be completed early in Cycle 5.

**C5-02 (Infrastructure Study report)** This is a VOTECH deliverable that was due last Cycle. It is part done and will be completed early in Cycle 5.

**C5-03 (VOTECH Science Framework Document)** This is a VOTECH deliverable that was due last Cycle. It is part done and will be completed during Cycle 5.

**C5-04 (Reduced Science Workshop Programme)** Our programme of science workshops has been extremely valuable, both in promoting understanding and use of AstroGrid, and in getting feedback on real science use of AstroGrid, enabling us to improve the system. They will continue to be of value in 2007, but to make room for our key priorities in this cycle we have cut the number back by half compared to the previous cycle.

**C5-05 (VOTECH 2nd baseline s/w release).** This is very simple - a web-page download centre making available prototype software written by all partners during the previous six months. We have been investigating a Gforge development site, but do not expect to implement this until at least the next cycle.

**C5-06 (Review/Refactor I&R process).** Code maturity is such that the existing I&R processes are no longer needed or appropriate. This review will consider alternatives and develop a proposal for I&R processes for the remainder of AG2 and beyond. It is anticipated that significant streamlining can be achieved.

**C5-07 (Deployed System Release 2007.1)** The next significant deployed system release will be timed to be ready for NAM.. This will be both a new end-user deployed system release, and a release of infrastructure software for use by developers, data centres, and other VO projects.

**C5-08 (VOTECH Stage-5 goals)** As usual, in AstroGrid internal planning, we will strive to align AstroGrid and VOTECH work, with the key method being to pro-actively promote VOTECH goals that fit our programme of work. The DS planning meeting is Mar 12-16, but the final TAP/Board approval is Apr 04.

**C5-09 and 15 (Tools Calls).** The AGSAG and the Management Team will review the outcome of the first Tools Call, with a particular view to whether the results produced were useful, and how much scientist and developer time was consumed. If this review is positive, we will proceed with a second Tools call, which will then use effort in Cycle 6 but not Cycle 5.

**C5-10 (NAM07)** As usual we will run a stall demonstrating AstroGrid software on a rolling basis. We have also made a request for a "science with AstroGrid" session.

**C5-11 (Component Configuration and Administration Mechanisms)** As well as being easy to install, AstroGrid services need to be straightforward to configure and administer. This is not to say the requirement for prior knowledge of the services can be eliminated, but that the mechanisms used to configure and administer the services should be consistent and unambiguous. C5-11 will produce such mechanisms.

**C5-12 (Documentation for science users).** The framework and structure is in place via our Plone-based web site, linked to help buttons on workbench components, and a significant fraction has been populated. The remaining sections will be produced in staged sections during the Cycle. As well as per-component pages, there will be a series of tutorials and recipes. For example, rather than writing an all singing all dancing "crossmatch" tool, various kinds of crossmatch related jobs can be done efficiently with AstroGrid workflows; what the user needs is a cookbook with recipes for such tasks.

**C5-13 (Fitness review).** Experience at workshops shows that users are quickly convinced that the logic of the VO in general, and the AstroGrid infrastructure in particular, is correct, and the right way to work in the future. They are also impressed with the user interface, but so far only a limited number of users are returning to AstroGrid for daily use. This is clearly partly because of a lack system robustness, which will be fixed during this cycle. However, there is a widespread feeling that in some places the tools don't *quite* do what people want or need, and that a critical review of "fitness for purpose" will lead to a much improved experience. This does not mean the shallow surface - arrangement of buttons, windows, etc - or the deep interior - basic relationship of components - but something in between - the way that chunks of interface are organised. Several ideas are floating around already. In addition there is a perceived need to construct an on-line community for AstroGrid users - ways for them to keep in touch and exchange experience and so

on. During this Cycle we do *not* intend to overhaul the UI; rather we will complete a study and make recommendations. If required changes look simple, they will be implemented during Cycle-6; if they are more radical or extensive, they will be implemented during the initial year of the planned UK VO service, for its first upgrade.

**C5-14 (Documentation for Technical Users).** Documentation for technical users and installers of the AstroGrid software components (tools writers; data centre staff; other VO projects; departmental system managers), like the documentation for science users, is underway but not complete. Completing this documentation is a high priority for this cycle. The category of tools writers is at the interface between science users and technical users.

**C5-16 (IVOA commitments).** We have a detailed plan for progressing standards work, with the key date being the interoperability workshop in Beijing in May. These are the key elements :

VOQL : TAP Telecons/emails; Integrate TAP & new ADQL into DSA; Track/direct ADQL standard

UWS: early draft document and implementation for May Interop

SSO: early document for May Interop

VOSpace : Complete and agree V1,0 specification; V1.1 specification for May Interop; V2.0 draft for discussion at May Interop

Registry : Progress V1.0 specification at May Interop; Test interoperability with ESO/NVO registries; Work on metadata curation; Progress ST(ime)AP protocol

AstroScope : Support Vizier in AstroScope / AR; Expose AstroScope engine

Applications WG : Help form/chair group; Track VOClient; Prepare PLASTIC IVOA note

**C5-17 (System performance).** The top priority this cycle is to improve software performance (speed), scalability, robustness, and reliability. A complete test and improvement programme is planned which will complete by the end of the cycle. In a little more detail :

Performance: Eliminate bottlenecks and poorly performing components. VO services should be perceived to cost no more than using existing solutions solving problems of similar complexity, noting that VO services are aimed at solving complex, high data volume and processor intensive problems.

Scalability: Verify System performance can be maintained by deploying additional VO services. Aim to support 100 logged in users, 20 simultaneous users of core services

Robustness: Review and enhance Unit and Integration tests, ensure components pass tests.

Document resource requirements by component. Verify components fail gracefully when resources squeezed.

Reliability: Ensure components report true error conditions and fail gracefully. Provide additional APIs in AR to support retry and fail-over. Components should try to avoid hanging where possible.

**C5-18 (Taverna prototype)** Following initial investigations, C5-18 seeks to prototype use of Taverna within the VO environment. In the first instance this means configuring Taverna to use astronomical extensions and eliminating the Life Sciences aspects. Initially we will use Taverna V1.5 which we believe has the framework we need for this work. From the prototype we will develop a plan for production use with the eventual aim of retiring the Workflow Builder UI component and the JES service.

**C5-19 (Applications Wish List).** What we refer to here is not general purpose UI driven tools such as Aladin or TopCat or VOSpec, but specific text-driven science applications common in the community, such as photo-ionisation codes, and photometric redshift codes. In the long term developers of such applications will understand themselves how to wrap them and offer them as VO services. In the meanwhile, we will wrap them ourselves and run them from our own AstroGrid

servers, offered as a list of "Tasks" in the Task Launcher. From our workshops, a fairly clear wish list of such applications has arisen. (The list can be found on the wiki [here](#).) Some of these are much easier than others. We will work through the list in order of ease-of-win.

**C5-20 (Convert key workflows to Python scripts).** The general concept of workflow has been well understood and liked by users, but the specific implementation is unsatisfactory, which currently wastes one of AstroGrid's key strengths. This is being fixed in two ways. Firstly, we are adopting Taverna, which is a superior user interface to our current workflow builder, and is much easier to use. Secondly, we will promote the use of scripting to construct very flexible workflows, which our Astro Runtime makes possible, and which is what more advanced users will want. Users can in principle use any script which uses XML-RPC, but Python is our initial target, having already constructed a Python library of VO service calls.. Users can already write such scripts, but it is not widely understood or appreciated, so we aim both to include scripting recipes in the documentation, and to provide existing workflow examples as Python scripts.

**C5-21 (Resource Browser).** So far our Registry browser and Task Launcher tools are straightforward text-searches on the Registry. We have long recognised the need for a more structured browser approach for users to simply "find out whats there". This tool is partway done and will be completed in this cycle.

**C5-22 (STILTS wrapping).** STILTS is the library that underlies the very successful TOPCAT tool. Once we complete this wrapping, the components can be called within an AstroGrid workflow, for example to construct a very flexible multi-catalogue crossmatch.

**C5-23 (Other component developments)** . Along with the major deliverables listed above, we plan to make minor upgrades or enhancements to many components. We will work as far as possible through the list below, if the higher priority deliverables allow. Work is planned on :

DSA ; Security ; CEA ; Jetty; UWS-PA ; Condor/EGEE/NGS integration ; VOspace prototype Registry V1.0 upgrade ; ST(ime)AP service ; ADQL Parser / Query Builder ; AR enhancements Leicester service upgrades [*Web/CVS/etc running on VMs*]

**C5-24 (Develop key external collaborations)** . In the last two cycles, we became increasingly successful in developing constructive collaborations with other projects. During this cycle we intend to push these further forward. These collaborations take considerable time and effort, but are crucial to ensuring the long term viability of AstroGrid software. Key collaborations are as follows :

CACR: NESSSI [*NVO Extensible, Scalable, Secure Service Infrastructure*) enables secure, asynchronous services to run on a Grid cluster]. Work with CACR to produce a CEA compliant interface to NESSSI for the injection of approved jobs into e.g. Teragrid

IPAC: Montage [*a toolkit for assembling FITS images into custom mosaics*] . Work with IPAC to produce a CEA service for Montage. Deploy CEA service on IPAC production systems

SDSC: SRB [*Storage Resource Broker – supports shared collections that can be distributed across multiple organizations and heterogeneous storage systems*] . Develop SRB capability within VOspace (using SRB JARGON toolkit) allowing the storage of data within file systems or SRB as appropriate

ESAC: DAL toolkit. The ability for users to publish their own datasets, FITS files etc as service within the VO has long been an unfulfilled target of AstroGrid. Our colleagues at ESAC however have developed a "Data Access Layer Toolkit", and are keen to integrate this with the AstroGrid DSA component. Working with them will therefore satisfyingly accomplish this long term goal.

CADC: Investigate collaboration opportunities

ChinaVO: Assist in deployment of a full AstroGrid service suite under the control of the ChinaVO

## (7) Cycle-4 deliverables for reference

### Software Product

C4-01 Complete external component review	2006 Sept 01
C4-02 Agree key standards at Moscow IVOA	2006 Sep 20
C4-03 Single point failure review	2006 Oct 01
C4-04 New cross-match functionality	2006 Oct 15
C4-05 New component admin and config methods	2006 Nov 01
C4-06 Review and improve component performance	2006 Nov 15
C4-07 VOTECH Data Exploration Study report (VOTECH)	2006 Dec 20

### Science Services

C4-08 ADQL service to UKIDSS, SDSS, 2MASS	2006 Sep 15
C4-09 Data access deployment plan for rest of project	2006 Sep 25
C4-10 Access to more key data sets (see notes)	2006 Oct 01
C4-11 Science Framework Document (VOTECH)	2006 Oct 15
C4-12 Expanded applications library (see notes)	2006 Nov 15
C4-13 Deliver programmes from Science Tools Call No. 1	2006 Dec 15

### Community

C4-14 Demo at IAU General Assembly, Prague	2006 Aug 14
C4-15 Set VOTECH Stage-4 goals	2006 Aug 25
C4-16 Developer/Data Centre Workshop(s)	2006 Sep 15+
C4-17 Complete web page re-factoring	2006 Sep 25
C4-18 Further Science Workshop(s)	2006 Oct 01
C4-19 VOTECH revised project plan	2006 Oct 15
C4-20 Deployed system release 2006.4	2006 Nov 15