The AstroGrid MySpace System

A.C. Davenhall,

Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK.

C.L. Qin, K.T. Noddle,

Department of Physics and Astronomy, University of Leicester, University Road, Leicester, LE1 7RH, UK.

N.A. Walton,

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK.

Abstract. MySpace is a component of AstroGrid, the Virtual Observatory system being developed in the UK. It provides AstroGrid users with work space for both temporary and long-term storage. Typically the system will be used to store datasets generated by queries submitted to large databases. The novel feature of the MySpace system is that although the work space is geographically dispersed, with stores at the various sites hosting AstroGrid archives, the user can access and navigate it seamlessly and easily, with the network details of the individual stores being hidden. MySpace is a fully integrated component of the AstroGrid system, written in Java and communicating via Web services. It is under active development and its current state and future plans are described. Functionality similar to that of MySpace seems likely to be a common requirement for a Virtual Observatory system, and the experience gained with MySpace should be applicable elsewhere.

1. Introduction

AstroGrid[1] is a UK e-Science initiative to develop a complete Virtual Observatory infrastructure (see Lawrence 2003 and other papers in these proceedings). It will provide seamless remote access to archives and databases. AstroGrid is the UK contribution to the global effort to develop a Virtual Observatory and it participates in the IVOA (International Virtual Observatory Alliance)[2] initiative. In particular it is developing and demonstrating prototype infrastructure of the sort which will be required by the Virtual Observatory.

---

[1] http://www.astrogrid.org/

[2] http://www.ivoa.net

Accessing astronomical archives basically consists of performing database searches to yield files of results. The astronomer may wish to download such files to his own computer for further analysis, or to perform further queries on the results, either in isolation or in combination with searches of other archives. In any event, work space is required for both temporary and long-term storage of such files, and software is needed for managing and accessing them. MySpace is AstroGrid's system for this purpose. Its novel feature is that it provides the astronomer with seamless access to geographically dispersed work space, typically with storage available close to the various archives, and hides all the details of accessing these remote systems. In addition to storing a user's results, MySpace is used by various components of the AstroGrid system as a convenient mechanism for storing temporary files. Qin et al. (2003) have described an earlier version of MySpace in slightly greater technical detail.

2. Design

The AstroGrid software will eventually be deployed at data centres and similar institutions providing access to data archives in the UK. This deployment will include several MySpace systems, providing work space for a variety of different purposes. There are two contrasting configurations which are likely to be particularly important: 'Cache systems' and 'Community systems'. Cache systems are large caches close to the data archives intended for the storage of large, transient files generated from the archives, whereas Community systems are longer-term, personal work space where astronomers can keep results and 'work-in-progress'. However, all MySpace systems have the same basic structure, irrespective of how they are being used (see Figure 1). Every MySpace system comprises: one 'MySpace Manager' and one or more 'MySpace Servers'. All of these components can optionally be geographically dispersed from each other and from other components of the AstroGrid system.

MySpace Manager embodies the intelligence of the MySpace system and is invoked by external components to access or manipulate files in MySpace. It incorporates an internal 'registry' comprising a list of all the files in the MySpace system.

MySpace Server is a repository where files are kept. It is invoked by the manager to copy, delete *etc.* a specified file and has little local intelligence.

An additional component is the 'MySpace Explorer' which allows an astronomer to interactively browse his files in MySpace. The user sees files in a hierarchical structure, with 'containers' as analogues of directories. This notional hierarchy of containers is distributed across (geographically dispersed) servers. The container hierarchy does not correspond to the actual directory structure on the servers (which is flat). The hierarchy is stored in the MySpace registry and its integrity is maintained by the MySpace Manager.

Though the three components of MySpace (Manager, Server and Explorer) are intended to work together, they can be installed separately. They communicate remotely and do not need to be co-located. This arrangement allows, for example, a Manager to control servers at remote sites. MySpace interacts
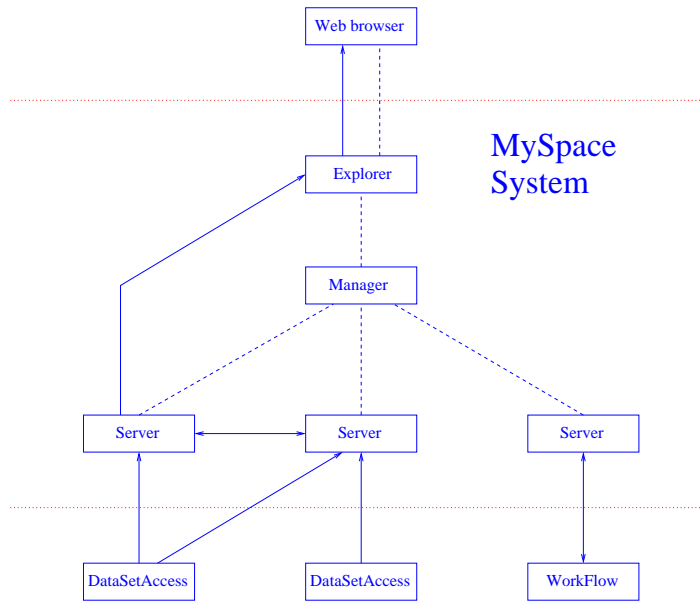
Figure 1.   Schematic diagram of the MySpace system. Broadly, the
dashed lines show the flow of commands and the solid lines the trans-
fer of files. DataSetAccess is the AstroGrid component for accessing
archives and databases. A MySpace Server will usually be co-located
with a DataSetAccess component for efficiency, but the architecture
does not require this. A number of AstroGrid components use MyS-
pace as a mechanism for temporary storage and WorkFlow is shown as
an example. All the external components communicate with a MyS-
pace system via commands sent to the Manager, in the same fashion
as the Explorer, but these flows of commands are omitted for clarity.

with other AstroGrid components to ensure that access is secure and controlled.
Nonetheless, a MySpace system can, in principle, be deployed in isolation or
with external Virtual Observatory components. MySpace is primarily intended
to provide work space and thus an optional mechanism is provided to retire old
files after a configurable 'expiry period' has elapsed.

3.   Technology Choices

MySpace is a component of AstroGrid and is being implemented using exactly
the same technologies as the rest of the system. It is written in Java and its
components communicate via Web services (both internal components within a
MySpace system as well as MySpace communicating with external components).
Insofar as practical, only open-source packages are used. Web services are im-
plemented using Axis[3] from the Apache Software Foundation. The AstroGrid

---

[3]`http://ws.apache.org/axis/`

system is accessed from a standard Web browser via a portal implemented using Cocoon[4]. The MySpace Explorer is part of this portal.

4.   Current Implementation

AstroGrid is being developed in a sequence of three month iterations. The version of MySpace available after the completion of Iteration 3 at the end of September 2003 included the following functionality: create a container, import a file, copy a file, move a file, change the owner of a file, advance the expiry date of a file, export a file, delete a file or container, look up details of a single file, look up details of all files matching an optionally wild-carded file name, look up a list of expired files, create a new user, delete a user and a rudimentary MySpace Explorer. Though much of the envisaged functionality is missing or incomplete the basic framework is present.

5.   Enhancements and Open Issues

During the remaining iterations of AstroGrid additional functionality will be added to MySpace. Some of the enhancements are conceptually straightforward and merely involve implementing functions which are not yet written. Similarly, the MySpace registry is currently implemented as a flat file and we plan to replace it with a set of database tables. Conversely, some of the desired features pose more substantial problems, which require further thought. These open issues include the following items.

Currently the MySpace system stores files. However, AstroGrid results are usually generated from database queries which can be represented as new database tables. Thus, we hope to include temporary database tables in MySpace. The current access control to files in MySpace is rudimentary. We will tie the access control for files to AstroGrid's wider system for controlling access to resources. There are several protocols which could be used for transferring files into and out of MySpace (HTTP, FTP, SFTP, GridFTP *etc*). We plan to support several of the common protocols, so that the most suitable available at a given site can be chosen. Finally, users may wish to 'publish' completed work by making files of results available to other AstroGrid users, so that they can be queried like public databases. Though challenging, none of these problems appear insuperable and we plan to address them in future iterations.

References

Lawrence, A. 2003, in Proceedings of UK e-Science All Hands Meeting, Nottingham, ed. S. J. Cox (Swindon: EPSRC), 428

Qin, C. L., Davenhall, A. C., Noddle K. T. & Walton, N. A. 2003 in ibid, 361

---

[4]`http://cocoon.apache.org/`