

# **Assessment of How the ASTROGRID System May Need to be Modified to use Global Geodata**

Jim Hogg  
Graticule Ltd (j.hogg@graticule.com)

## **1. Introduction**

The AstroGrid system provides a significant set of software tools for astronomers and other scientists for accessing and using a range of astronomical functions and archived astronomical data for research. Spectra, images, tables, videos, statistics, maps, graphs and other types of data can be retrieved and analysed from a workbench installed on a computer with an Internet link. Typically, data are archived in a distributed global archive system and processing of data takes place on a global array of remote virtual private servers or clouds. Results can be archived in a cloud, processed in a cloud and, if necessary, viewed on a workbench display screen, depending on the type of results.

Many of the generic functions in AstroGrid are already being used by global environmental research scientists for viewing, interpreting, measuring, monitoring and assessing changes over time in global forest types and cover, global ecological research and estimating amounts of carbon released or sequestered in the earth's crust involving relatively large sets of sequential satellite global geodata of the earth's surface.

The author of this report has not had the benefit of training and use of the AstroGrid system. Without theoretical knowledge of its huge range of software packages, functions and integrated workflow tools and practical experience of using at least a selection of the software tools available, the author does not feel able to provide a balanced, critical and realistic assessment of the potential of the AstroGrid system for applications involving geodata. Moreover, although the author has knowledge and experience of satellite remote sensing systems, digital image processing and classification procedures, he has limited knowledge and experience of measuring and monitoring the extent of global or regional forest cover and change, global ecological research and carbon release and sequestration research. So the author does not feel able to provide explicit recommendations and a rigorous assessment of how AstroGrid may need to be modified for use of global geodata.

The aims of this report are therefore to highlight some of the features, software tools and networking capabilities of the AstroGrid system and discuss what types of modification, if any, may need to be made before the AstroGrid system would meet basic requirements for processing global environmental data.

The structure of the document is in three parts. The first covers the parts of the AstroGrid system which may require minor modifications but would probably work as they are. The second part covers the parts of the AstroGrid system which do not appear to require significant change as the software functions are generic and widely applicable. And the third part covers the formats of data which the AstroGrid system appears to include for handling global environmental geodata.

## **2. Minor Modifications to the AstroGrid System**

A user must register online to obtain a personal user name, download the AstroGrid Workbench and obtain a personal Workspace area. Registration provides access to the Registry which enables a user to search for and use predominantly astronomical data but some other types of data as well. Although the search system already works with different types of data, it may require the addition of facilities for geographical searches such as use of a gazetteer of place names, regions, countries and other phenomena and descriptive names of types of forests, types of ecological habitats and other features. AstroGrid already has the earth's geographical reference systems, a plethora of map projections and different models of the shape of the earth. Graphical grids representing many different 2-D and 3-D map projections can be superimposed on images of the earth taken from earth orbital satellites and other platforms.

The Registry enables a user to search for images, names of places, areas defined using world longitude and latitude co-ordinates, catalogues and other environmental features provided they are already in the catalogues. When a user searches for particular types of data covering a specified geographical area, the Registry returns the archives a user sought. The search queries are typically sent to where the archives stored and are retrieved and stored on a user's MySpace or local disc. A user can then graphically display all or parts of the data on his workbench and submit them for processing in the cloud.

The Application Launcher has been designed to meet the requirements of astronomers but it may be adapted to handle geographical data of the earth. The software tools in the Application Launcher should work in an analogous way to how it works with astronomical data but may need some modification depending on the types of data involved and the requirements of a researcher.

Similarly, the Query Builder should be capable of working with data about the earth but may need to be adapted to meet the specific requirements of earth scientists. In order to use the Query Builder to perform searches of geographical data, it will require a geographical data dictionary rather than an astronomical data dictionary. The Query Builder itself would probably be capable of being used as it is. Of course, if forest cover research scientists have a particularly demanding set of queries, the Query Builder will have to have access to the relevant databases and data in order to perform its task efficiently.

The AstroScope tool for finding data over a region of the earth should work with catalogues of geographical data about the earth, its geographical features, extents, elevations and other variables. This appears to perform in a similar way to the search features provided by the United States Geological Survey.

Application Launcher should work with geographical data although only with a relatively small set of functions. Some functions will no doubt have to be adapted to meet the specific requirements of forestry, ecology and carbon research scientists.

QueryBuilder is a pre-packaged workflow system to carry out common procedures. The system itself should be capable of working with global functions such as map projections of the earth but will require data dictionaries for each research field such as global forest cover, types and extents.

Upload and Save data in your personal workspace area should work providing the data is in one of the available formats accepted by AstroGrid.

From the above summary of selected tools, work will have to be done to adapt some of the software tools for use with the types and formats of global forest data. In some cases, this will only require the substitution of a global forest data dictionary with locations, elevations, aspect, slopes and forest types, age, average height, average density, volume of standing timber and such like.

### **3. Generic Software tools in the AstroGrid System**

The large collection of software packages appears to provide a massive suite of generic software tools. Many of these are specific tools for astronomical research as one would expect but there are among these many tools which could potentially benefit scientists working with global sets of geodata about the earth.

The generic suites of software tools provide a rich source of standard functions which are widely application in environmental research. Rather than spend a great deal of time designing and building new software tools for specific projects involving global forests, ecological research and carbon studies, the AstroGrid system offers an instant solution to the provision of custom software for research. By using workflow software that forms an integral part of Astrogrid, specific novel software applications can relatively quickly and easily be built by specifying functions and sets of data to be processed in a workflow sequence.

The challenge for research scientists in global forests, ecological dynamics and carbon research lies in learning and training in how to make effective use of the AstroGrid system. An experienced computer scientist should be able to understand the AstroGrid system and make effective use of it for research but it is necessary for global forest, ecology and carbon scientists to at least become acquainted with the potential and scope for using the system in their research.

The most effective way of stimulating curiosity and interest of the target group of scientists working in global forest change, ecological change and carbon sequestration would be to hold a series of workshops to reveal what global research scientists using geodata could achieve by using the AstroGrid system. The workshops would probably have to be progressive with lectures, exercise schedules and practical laboratory training on how to use selected relevant parts of the system. For example, a set of scripts for reading a file of geodata, processing it through selected functions, creating a map, graph or table of statistics and archiving the results in a relational database would in my view stimulate the interest of research scientists.

Just as the AstroGrid system enables undergraduate students in astronomy to perform research by accessing software functions and astronomical data archives from around the world on a workbench in a university, research scientists in global forest research should be able to undertake research using chosen generic software tools and archived global geodata stored around the world.

Training should in my opinion focus on the scope and potential of the AstroGrid system for global forest research. It should enable scientists to run a prepared Java script or similar script to initiate the processing of chosen sets of global data. It should produce results in different forms such as images, maps, graphs and videos. The focus of the training should include the critically important interpretation, analysis and conclusions to be drawn from the results. In short, the training should demonstrate that scientists can easily and quickly run such a script, edit it and rerun it to process different types of global geodata for research purposes.

#### **4. Formats of Data**

The AstroGrid system has functions for reading and writing digital image data in many different formats. These include images in FITS or JPEG formats. Aladin supports FITS, JPEG, PNG and GIF images.

First standardised in 1981, the Flexible Image Transport System (FITS) is the data format widely used in astronomy for transporting, analysing and archiving scientific data files. FITS is much more than just another file format such as JPG or GIF. It was primarily designed to store scientific data sets consisting of multi-dimensional arrays (images) and 2 dimensional tables organised into rows and columns of information. These are in essence matrices.

FITS version 3.0 was standardised in 2008. FITS was designed with an eye on long-term archival storage. The maxim 'once FITS always FITS' represents the requirement that developments to the format must be backwards compatible.

The most common type of FITS data is an image header/data block. The term image is somewhat loosely applied. The format supports data arrays of arbitrary dimension – normal image data are usually 2-D or 3-D. The data themselves can be in one of several integer or floating point formats, specified in the header.

FITS image headers can contain one or more scientific co-ordinate systems that are overlaid on the image itself. Images contain an implicit Cartesian Coordinate System that describes the location of each pixel in the image but scientific uses usually require working coordinates for example the celestial coordinate system. As FITS has been generalised from its original form, the world co-ordinate system has become more and more sophisticated: early FITS images allowed a simple scaling factor to represent the size of pixels but recent versions of the standard permit multiple non-linear coordinate systems, representing arbitrary distortions of the image. The world coordinate system standard includes many different spherical projections.

FITS also support tabular data with named columns and multi-dimensional rows. Both binary and ASCII text table formats have been specified. The data in each column of the table can be

in a different format from the others. Together with the capability to string multiple headers/data together, this allows FITS files to represent entire relational databases.

The FITS model supports a wide range of image, graphic, database, video and other formats. It is increasingly being used in open source geographical information systems, NASA, the US Geological Survey and other Government organisations around the world. Vector point, line, polygon and text data can be archived in a relational database, accessed on demand, processed and stored back in an archival relational database.