

# ASTROTROP PROJECT

## Draft Executive Summary of Final Report

### 1. Introduction

The main aim of the ASTROTROP project is to evaluate the potential to adapt AstroGrid/Euro-VO virtual observatory software so it can be used by members of the TROPGLOBE network for virtual overlays of tropical forest datasets in a virtual Pan-Tropical Forest Observatory. This discussion document reports the findings of the evaluation exercise, and suggests the next steps which TROPGLOBE members could take to demonstrate the concept practically, and agree on the standards needed to make a Pan-Tropical Forest Observatory operational. It is based on the work of Dave Morris, Keith Noddle and Andy Lawrence at the Royal Observatory, Edinburgh, which they reported at a workshop in Leeds earlier in 2015. It is being circulated as a basis for consultation and discussion in advance of the Second ASTROTROP Conference, in February 2016, which will discuss the issues identified here.

### 2. Background

Since remote sensing satellites have been collecting global data since 1972, ideally these data would have been converted into global digital information on the distribution of forests, other ecosystems and land use which scientists could analyse. Instead, we lack regular measurements of changes in world forest area, and global environmental change scientists have had to rely on compilations of unreliable national statistics. This has limited the accuracy of scientific research into global environmental change, and of information available to policy makers.

A World Forest Observatory would produce annual maps of world forest area, biodiversity and carbon density which scientists worldwide could analyse. The initial plan for a World Forest Observatory in 2010 relied on a *hub* design, in which 10-20 teams across the world would produce digital maps of three forest attributes; area, biodiversity and carbon. These maps would then be combined on a central geographical information system at a global hub, from which users would download global digital maps. However, in an alternative *virtual* design, the digital maps of multiple forest attributes would remain on the information bases of the teams which produce them, but sophisticated software would enable the multiple attribute maps to be overlaid in geo-referenced format in the computers of users which download them.

Establishing a Pan-Tropical Forest Observatory in the UK as a pilot for a World Forest Observatory is highly appropriate since UK scientists are leaders in tropical forest research. It would also act as a focus for a vibrant industry devoted to developing the innovative technologies needed for a growing family of global

environmental observatories that can ensure a seamless chain from satellite data to information on policymakers' desks, consistent with the goals of the UK Satellite Applications Catapult, and of the European Commission's Copernicus Programme.

The ASTROTROP project has brought together astronomers, who are members of the AstroGrid/Euro-VO network, and tropical forest scientists, who are members of the TROPLOBE network. It is funded by the Science and Technology Facilities Council's Challenge-Led Applied Systems Programme (CLASP), which aims to: apply STFC-Funded Research to have impacts in STFC Global Challenge Areas, which include the environment; fill technology gaps, which include automated monitoring of ecosystems and carbon emissions; produce "demonstrable deliverables for a potential market", including scientists, companies, policy makers and other stakeholders; and have environmental, social and commercial benefits.

### **3. How the AstroGrid/Euro-VO Virtual Observatory Works**

A virtual observatory can be summed up as having "all the world's databases inside your PC", just as the Worldwide Web means that "all the world's documents are inside your PC". Users can search for data and services, and then combine and overlay data, images, catalogues, spectra and time series. Discovering the data that you need is facilitated by the metadata attached to each parcel of data, which describe, for example, wavelength, location and pixel size.

AstroGrid virtual observatory software has been refined through European partnerships under the umbrella of the International Virtual Observatory Alliance of astronomers (IVOA). It includes processes, documents, data, services and code:

- i. *Processes* include twice yearly international meetings of the standards body, and working groups and interest groups which agree on standards documents. Processes are modelled on the Worldwide Web Consortium (W3C), and involve both the IVOA and the multidisciplinary Research Data Alliance (RDA). Standards agreed by the standards body are normally ratified by the IVOA generally.
- ii. *Documents* include: data models which define how data are structured, and standards for expressing location in space and time, registry metadata, service protocols and data query language.
- iii. *Data* include: registries (or Yellow Pages), collections of images and spectra, table-like databases, and events.
- iv. *Services* include registry querying, image and spectrum grabbing, database querying, and event querying. After data sources are identified in registries other services are used to grab data from identified locations.
- v. *Code* includes user tools, such as Topcat and Aladin; code for running services; and libraries.

Consequently, the astronomical virtual observatory is not mainly about software, but the result of a huge community effort by dozens of people over 13 years to agree on the standards needed to make it possible. Standards are important for applications,

services and data exchange. If tropical forest scientists wish to replicate AstroGrid/Euro-VO then they need to replicate the communal efforts of astronomers.

The Working Groups and Interest Groups produce standards. The IVOA has Working Groups on Applications; Semantics; Working Groups: Data Access Layer; VO Event; Data Modelling; Resource Registry; Grid and Web Services; and VOTable. It has Interest Groups on: Theory; Education; Operations; Data Curation and Preservation; and Knowledge Discovery in Databases. Other Groups and Committees include the Standing Committee on Standards & Processes and the Standing Committee on Science Priorities (<http://wiki.ivoa.net/twiki/bin/view/IVOA>).

Virtual observatory standards evolved gradually, with simple things being tackled first, and members of the TROPLOBE network could emulate this approach. This would enable data and information sharing in the network to become operational as quickly as possible, and make the Pan-Tropical Forest Observatory operational too.

#### **4. Main Findings of the AstroGrid/Euro-VO Evaluation Report**

The needs of global change scientists for spatial information on tropical forests are similar to those of astronomers, which is why the case for refining AstroGrid/Euro-VO software for use in a Pan-Tropical Forest Observatory originally seemed so compelling. However, the main conclusions of the evaluation are that:

- i. Whereas there are a lot of large astronomical datasets, which can be accessed through a virtual observatory because their hosts are members of the IVOA, the same is not true for global change science. The data in most of the large datasets, such as those of NASA, NOAA, ESA etc., are not directly usable by global change scientists without intermediate processing into geospatial information. For the most part, geospatial information needed by global change scientists is held by individual researchers and relatively small research groups and networks.
- ii. Information required by TROPLOBE partners falls into two main categories: (a) information within the TROPLOBE network, which can be obtained by controlled access, while testing the use of metadata and publishing services; and (b) information outside the TROPLOBE network, which can be obtained by establishing links with the hosts of the requisite databases.
- iii. AstroGrid/Euro-VO software and standards are not reusable without significant modification and time because they are too specific to the needs of astronomers.
- iv. The structure of the AstroGrid/Euro-VO project is, however, reusable if TROPLOBE partners have the collective will to replicate the efforts made by astronomers in making a virtual observatory feasible, e.g. by establishing processes to agree on standards and describe these in documents, making their public datasets consistent with these standards, and writing code for data/information sharing.
- iv. Nevertheless, because other members of the geospatial and environmental science communities have been pursuing the open data concept for some time now, sufficient open source geospatial standards and software exist to form the basis for

prototyping a Pan-Tropical Forest Observatory. Table 1 shows the geospatial equivalents of the components of AstroGrid/Euro-VO software.

Table 1. Geospatial equivalents of the Components of AstroGrid/Euro-VO software.

	Astronomy	Geospatial Equivalents
i. <i>Data models</i>	Obscore	Worldfile GIS metadata Ecological Metadata language
ii. <i>Standards bodies</i>	IVOA	Open Geospatial Consortium
iii. <i>Registries</i>	AstroGrid/Euro-VO Registry	Comprehensive Knowledge Archive Network (CKAN)
iv. <i>Registry and Data Grabbing Services</i>	SIAP	Geoserver
v. <i>Data Manipulation Tools</i>	Aladin	QGIS

CKAN is used by Open Data Germany ([http:// open-data.fokus.fraunhofer.de/en/ernten-und-geerntet-werden-erfahrungen-beim-govdatadeharvesting/](http://open-data.fokus.fraunhofer.de/en/ernten-und-geerntet-werden-erfahrungen-beim-govdatadeharvesting/)) and by the US National Geothermal Data System ([www.geothermal-energy.org/pdf/IGAstandard/SGW/2013/Clark.pdf](http://www.geothermal-energy.org/pdf/IGAstandard/SGW/2013/Clark.pdf)). Other relevant geospatial initiatives include: (i) the Global Index of Vegetation-Plot Databases (<http://www.givd.info>), a registry containing metadata for 226 vegetation plot databases and links to them; (ii) the Data Observation Network for Earth (DataONE), a registry that aims to store, and give access to multi-scale, multi-discipline, and multi-national science data (<https://www.dataone.org/software-tools/tags/GIS>); and (iii) the Knowledge Network for Biocomplexity, a network of federated institutions within DataONE , that have agreed to share data and metadata using a common framework, principally revolving around using Ecological Metadata Language as a common language for describing ecological data (<http://www.dcc.ac.uk/resources/implementations/knb-knowledge-network-biocomplexity>).

## 5. A Blueprint for Prototyping a Pan-Tropical Forest Observatory

This section outlines the demonstration facility for prototyping a Pan-Tropical Forest Observatory which is being assembled on the ASTROTROP website by the team at the Royal Observatory, Edinburgh.

### 5.1 Registry Software

The first component is the Registry, which links Databases and connects them, in turn, to Users.

Two candidates for Registry Software are Metacat and CKAN. CKAN has been chosen because it is available now; does what the IVOA Registry does; has an adaptable metadata store; is widely used in open data services in the USA; and is being continually updated by a large active community. The University of Bristol, an

ASTROTROP partner, uses a CKAN service to structure its data repository: data.bris.ac.uk. Each partner could have its own repository which forms a sub-directory that is part of the central directory of an overall repository. Each dataset should be described by a metadata file that includes such elements as: (i) attribute; (ii) date created; (iii) date last updated; (iv) data type; (v) spatial coordinates.

### 5.2 *Registry and Data/Information Grabbing Services*

Once the locations of required data or information are identified they can be extracted from their host databases and channelled to a user's computer. The chosen candidate for Data/Information Grabbing Software is Geoserver. This can "cut out" from a large geospatial database the data/information for an area defined by latitude and longitude coordinates, and is consistent with GeoTIFF and CSW formats. CKAN itself does not handle GIS data but has a plug-in extension for Geoserver. Geoserver is consistent with six standards for data formats, so it should be suitable. Partners, and end users, might expect to see outputs in Computable Document Format (CDF) and in GeoTIFF format, which embeds georeferencing information.

### 5.3 *User Tools*

Once data have been channelled to a user's computer they can be analysed using standard GIS software tools, such as ArcINFO and R.

### 5.4 *Alternative Architectures*

Other architectures are possible too. For example, a few years ago Duncan Golicher of the University of Bournemouth proposed using GeoServer to channel open information layers to the computers of end-users, where they would be stored on PostGIS information bases and analysed using R software.

### 5.5 *Undertaking Demonstration Projects*

TROPGLOBE members are invited to use the facility available on the ASTROTROP website to undertake informal demonstrations of overlaying their data. The resulting overlays will be presented at the Second ASTROTROP Conference in February.

It would be helpful for many reasons for TROPGLOBE members to use a standard template to describe their overlays. This could include:

- i. Primary researcher
- ii. Science goal
- iii. Datasets
- iv. Problem description
- v. Current solution
- vi. VO solution
- vii. Spatial cover
- viii. Spatial extent
- ix. Formats for metadata, input and output information files
- x. Validation and errors protocols
- xi. Policy relevance

## **6. Consultation and Discussion on Next Steps**

### *6.1 Standards and Processes*

One of the main items on the agenda of the Second ASTROTROP Conference in February will be to agree on initial temporary standards and on a set of ongoing processes to generate permanent standards for exchanging and overlaying data. Alan will be consulting with as many members of TROPLOBE as possible over the next two months to get your thoughts on this and identify your needs.

### *6.2 Incubator Activities*

It is important to link ASTROTROP standards to those needed by other groups, including commercial and governmental groups, as well as international scientific groups such as ICSU. Jim Lynch has been exploring various possibilities, which will be discussed at a session at the Second ASTROTROP Conference.

### *6.3 Interface Activities*

This pioneering project has shown the benefits that can arise from collaboration between astronomers and global change scientists. Another session at the Second ASTROTROP Conference will explore other potential areas for collaboration, and we invite suggestions for this from members of the AstroGrid/Euro-VO network.

### *6.4 Follow Up Projects*

Various initiatives have been made over the last two years to gain funding for projects to follow up ASTROTROP by further developing virtual forest observatory technologies and exploiting their benefits. The University of Leicester wishes to exploit synergies between its successful GlobBiomass proposal and ASTROTROP technologies. There is also potential to participate in proposals to the EU Horizon 2020 European Research Infrastructures (including e-Infrastructures) Call. There are many other possibilities and all suggestions for these will be welcomed.