

ASTROTROP WORKSHOP

20-21 July, School of Geography, University of Leeds.

Minutes

Monday 20th July

1. Welcome and Introduction (AG)

Alan Grainger welcomed ASTROTROP partners to the University of Leeds, and to the new location of the School of Geography in the connected Garstang and Manton Buildings. The workshop was held in Seminar Room 2 in the Garstang Building.

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 brings 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.

The aims of this workshop were to discuss the evaluation of the potential to adapt AstroGrid/Euro-VO software so it can be used for virtual overlays of tropical forest datasets; and to refine the project plan for the remainder of the project.

2. AstroGrid/Euro-VO Fundamentals (AL/KN)

Andy Lawrence, supported by Keith Noddle, outlined the fundamentals of virtual observatory software. 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.

Current virtual observatory software is not merely AstroGrid, but has been refined through European partnerships under the wider umbrella of the International Virtual Observatory Alliance of astronomers (IVOA).

The virtual observatory 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 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 want 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. The first standard agreed was on VOTable, which enables the metadata of data stored in registries to be displayed in a tabular format. The next standards agreed were on Simple Image Access (SIA), which handles queries to registries, and Simple Cone Search, which defines the position in the sky for which data are required.

3. AstroGrid/Euro-VO Evaluation Report (KN)

Keith Noddle then summarized the findings of the AstroGrid/Euro-VO Evaluation Report. The needs of global change scientists for spatial information 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. AstroGrid/Euro-VO software and standards are not reusable without significant modification and time.
- ii. The structure of the AstroGrid/Euro-VO project is, however, reusable if TROPGLOBE partners have the collective will. Sufficient open source geospatial standards and software exist to form the basis for prototypes in the rest of the project.

Current astronomical and geospatial capabilities can be compared as follows:

	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, for example, by Open Data Germany ([http:// open-data.fokus.fraunhofer.de/en/ernten-und-geerntet-werden-erfahrungen-beim-govdata deharvesting/](http://open-data.fokus.fraunhofer.de/en/ernten-und-geerntet-werden-erfahrungen-beim-govdata-deharvesting/)) and by the US NATIONAL Geothermal Data System (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](http://www.dcc.ac.uk/resources/implementations/knb-knowledge-network-biocomplexity)).

The most fundamental difference between the astronomical and tropical cases is that 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.

Information required by TROPGLOBE partners falls into two main categories:

- i. Information within the TROPGLOBE network, which can be obtained by controlled access, while testing the use of metadata and publishing services.
- ii. Information outside the TROPGLOBE network, which can be obtained by establishing links with the hosts of the requisite databases.

4. Overview of Proposed Demo Software (DM/AL)

Building on the Evaluation Report's conclusions, Dave Morris proposed how to construct from existing geospatial software the Demo Software needed in the second stage of the ASTROTROP project to demonstrate the feasibility of taking a virtual observatory approach to a Pan-Tropical Forest Observatory.

4.1 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 many advantages since 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.

4.2 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.

One 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.

4.3 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.

4.4 Alternative Architecture

Duncan Golicher shared with the workshop the design for a similar system which he produced a few years ago. Open information layers would be channelled by GeoServer to the computers of end-users, where they would be stored on PostGIS information bases and analysed using R software.

4.5 Next Steps

The next step is for TROPGLOBE partners to identify a compact set of demonstration projects, the problems they are intended to solve, the types of data/information required for each project and the preferred formats in which they are stored.

5. Proposals for Demonstration Projects

Partners described the information sets which they can share and these were matched with one another through discussion.

5.1 Available Information Sets

- i. Leicester:
 - a. Mexico - Biomass map

- ii. Bournemouth:
 - a. Mexico - Plot data
 - b. Mexico - Forest area change maps for > 60 forest types

- iii. Bristol: Global ecosystem models of leaf area index, photosynthesis, respiration, carbon flux, biomass and burn area, run for 1950-2002 (with possible extensions to 1850 and 2015) for key biomes and 10 plant functional types with 50 km grid squares.

- iv. Leeds: carbon change in forest plots in Amazonia, Africa + Southeast Asia.
- v. Sheffield:
 - a. Mozambique and Guinea-Bissau - carbon
 - b. Southeast Asia - carbon
- vi. Edinburgh (Royal Botanical Garden):
 - a. Latin America - tree species in dry forest (point data and plot data)
 - b. Modelled biomes
- vii. King's College, London:
 - a. Latin America - forest area change map (Terra-i)

5.2 *Presentations on Information Sets*

Two of the information sets were the subjects of extended presentations.

5.2.1 Leicester

Kevin Tansey outlined the various digital map information sets which he is analysing or producing himself. They include three global forest carbon maps, from Saatchi et al. (2011), Baccini et al. (2012), and Santoro et al. (2013); four forest biomass maps for Mexico; and a beta version of a forest carbon map for the Congo Basin. By the end of the year Leicester will have produced a biomass map of Mexico for 2010 as part of the ESA GlobBiomass project. It was noted that there would also be a biomass map of Indonesia as part of the same project.

5.2.2 RAINFOR

Gabriela Lopez Gonzalez explained that RAINFOR started 20 years ago with 50 plots in Latin America; this has now risen to about 2,000 plots. RAINFOR is a network of friends working together and is expanding in scope to include dry forest data. Its companion AfriTRON network covers forests in Africa, where lianes are especially important. It has also initiated a global wood density database, called DRYAD.

Two websites cover forest plots throughout the world: ForestPlots.net, which includes over 1,000 plots; and the Global Index of Vegetation Plot Databases (GIVD) which, as its name implies, is a registry of 226 databases.

Initially, RAINFOR data were stored on an Access database. Because of difficulties in inputting data this was converted eight years ago into an Excel database. The reason for staying with the .xls format, rather than moving to the .csv format, is that this is preferred by the many South American scientists who work with RAINFOR.

One challenge faced by RAINFOR was to determine when, and how, to make its data public. It currently makes available data linked to published papers, and has worked with the British Library to devise a procedure to separate data from the papers using a DOI reference, so that the rights of those who collect data are still protected.

Although there are no standards yet for storing biomass data, Dave Morris suggested that because of the huge amount of effort devoted by RAINFOR partners to constructing this database its standard could be adopted as a defacto standard.

5.2 *Mutually Agreed Sharing Possibilities*

The following preliminary set of possibilities for sharing information in demonstration projects was assembled, together with the key science question:

- i. *Amazonia* carbon density (Leeds) - - Forest area change (King's)
Q: Is there any relationship between areas of forest cleared and carbon density?
- ii. *Mexico* Biomass (Leicester) -- Forest area change (Bournemouth)
Q: What is the rate of carbon emissions from deforestation in Mexico?
- iii. *Amazonia* Actual tree growth (Leeds) -- Tree species (Edinburgh)
Q: Is tree growth species dependent?
- iv. *Amazonia* Actual tree growth (Leeds) -- Modelled tree growth (Bristol)
- v. *Congo* tree growth (Leicester) -- Modelled tree growth (Bristol)
Q: Can measured tree growth validate modelled tree growth?
- vi. *Mexico* Dry forest species (Edinburgh) -- Tree Species (Bournemouth)
Q: What is the coincidence between the two species databases?

Tuesday 21st July

6. Next Steps in Demonstrator Projects (AG)

A representative selection of Demonstrator Projects is needed to show the feasibility of a Pantropical Forest Observatory.

6.1 *Selecting Demonstrator Projects*

At least five criteria could be useful for selecting these projects:

- i. Projects have a 'wow' factor that will impress participants at our closing Conference 2, since they are topical; show how a virtual observatory can solve a problem that other approaches cannot; and have a massive effect for not much work. Andy Lawrence gave an example from astronomy, published in *Science* (Chilingarian et al., 2009).
- ii. We choose exemplars of different types of information combinations.
- iii. The projects are feasible for partners to undertake, e.g. since they have digital information sets available to share, or have already established links with other partners (we can check on any informal exchanges in Year 1 of ASTROTROP).

iv. The projects are attractive to partners, e.g. because they bring different types of data/information together; each partner gains something from the collaboration; they tackle state-of-the art science problems; and they could lead to journal papers.

v. The projects have practical value for stakeholders, e.g. they could be useful in implementing the REDD+ scheme of the UN Framework Convention on Climate Change, or the RED List scheme of the World Conservation Union.

6.2 *Templates for Demonstrator Project Descriptions*

Keith Noddle suggested that it would be helpful for many reasons for partners to use a standard template to describe their Demonstrator Project proposals. He said that a template used by AstroGrid/Euro-VO could provide a useful starting point:

- i. Primary researcher
- ii. Science goal
- iii. Datasets
- iv. Problem description
- v. Current solution
- vi. VO solution.

Kevin Tansey showed a similar example from remote sensing science, which included additional elements that are particular to a global change study, including spatial cover; spatial extent; formats for metadata, input and output information files; validation and errors protocols; and policy relevance.

6.3 *Immediate Standards*

Dave Morris said that it would be helpful to specify standards immediately for the terms 'forest' and 'image'.

Forest is defined by FAO (2010) as "as land of at least 0.5 ha covered by trees higher than 5 m and with a canopy cover of more than 10%, or by trees able to reach these thresholds, and predominantly under forest land use."

Remote sensing scientists can make alternative estimates of the area of forest when forest is defined in terms of land with *tree canopy cover* of more than 25%, 50% or 75% (Hansen et al., 2013), though measurements for lower thresholds are difficult.

Images collected by satellites measure average canopy reflectance in pixels determined by sensor resolution. So unlike astronomical images, satellite images must be processed by remote sensing scientists into maps of forest area before scientists from other disciplines can use the latter. In global change science, therefore, satellite images constitute raw '*data*' whereas forest area maps constitute usable '*information*'.

6.4 *Miscellaneous Practical Issues*

It is important to identify: what additional resources are needed for the demo projects and what are already available; what additional software is needed and what is already

available; and what new standards are needed and what are already available. There was a brief discussion of these topics. Keith Noddle suggested that if we list the future actions needed on standards this will help partners who did not attend the workshop.

6.5 *Infrastructure*

Keith Noddle suggested that once the list of demonstrator projects has been identified, the information needed can be used to populate CKAN. He will set up the basic system in Edinburgh so that partners can use it to exchange information. Each partner should receive a copy of Geoserver to extract the information they need. 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.

6.6 *Website*

The project website needs to be updated with the Evaluation Report. Keith Noddle has set aside an internal area on the website for software and the minutes of meetings. Dave Morris proposed using the website to store the minutes of meetings on standards, the design of demonstrator projects etc., just as in the IVOA. He suggested holding 2-3 meetings for each demonstrator project. Project funds are available for these meetings.

7. Incubator Groups

Alan Grainger said that there is considerable potential for the virtual observatory software devised in this project to have practical commercial and policy uses, not least for the REDD+ scheme, even though we shall be using software different from that originally envisaged. A special incubator workshop could be held. Jim Lynch said that we are at a critical stage in negotiations on REDD+ in the lead up to the Conference of the Parties of the UN Framework Convention on Climate Change in Paris in December. This could unlock lots of funds for practical monitoring schemes, and the ASTROTROP project could even make a presentation at a CIFOR conference side event. News about our project could also be shared with other stakeholders at Conference 2 in February. As this will be in London we can invite insurance companies and risk assessment companies that attended a special set of NERC knowledge exchange workshops in the City of London a few years ago.

8. Interface Groups

In the original project plan it was intended to form interface groups at universities which had members of both the AstroGrid and TROPLOBE networks, i.e. Bristol, Edinburgh, Leeds and Leicester. Although a few meetings have been held we could maximize the impact of mutual learning between astronomers and global change scientists by holding a special workshop, perhaps at Leicester, at which astronomers could propose how their techniques could be used to study Planet Earth. The papers presented at the workshop could be published in a special issue of an appropriate journal. Andy Lawrence agreed to take soundings with astronomer colleagues.

9. Follow on Projects

9.1 *Horizon 2020 Projects*

Alan Grainger reported on two EU Horizon 2020 proposals submitted since Conference 1 in which he had been a partner and which included a virtual observatory component. They responded to calls for 'Making Earth Observation and Monitoring Data Usable for Ecosystem Modelling and Services' and 'Demonstrating the Concept of Citizen Observatories'. Both proposals, unfortunately, had been unsuccessful.

9.2 *ESA Projects*

Kevin Tansey, on the other hand, had been a partner in a successful proposal to the European Space Agency (ESA) for a GlobBiomass project (2015-17) which would use satellite data to provide improved quantification of regional biomass for 2005, 2010 and 2015 and global biomass for 2010. He saw great potential for synergies between this project and a virtual observatory devised by ASTROTROP, e.g. in overlaying maps from different satellites, overlaying these maps with maps based on plot data, and estimating the uncertainty of these maps.

9.3 *Standards Initiatives*

There was a general consensus that TROPGLOBE partners should engage with existing initiatives to devise standards for global terrestrial measurement. An immediate opportunity for funding such work is the EU Horizon 2020 European Research Infrastructures (including e-Infrastructures) Call, which will be announced in September. As AstroGrid/Euro-VO has shown, standards are an essential part of any e-infrastructure. Duncan suggested that collaborating in a European initiative would also be a useful stimulus for devising Mexican standards. Another related initiative is the ICSU World Data System (www.icus-wds.org).

References

- Baccini A. et al., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change* **2**, 182–185.
- Chilingarian I. et al., 2009. A population of compact elliptical galaxies detected with the virtual observatory. *Science* **326**, 1379-1382.
- FAO, 2010. *Global Forest Resources Assessment 2010*. FAO Forestry Paper No. 163. UN Food and Agriculture Organization, Rome.
- Hansen, M.C. et al., 2013. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850-853.
- Saatchi S. S. et al., 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences USA* **108**, 9899–9904.

Appendix 1

List of Participants

Pru Foster (Bristol): glpnf@bristol.ac.uk
Duncan Golicher (Bournemouth): dgolicher@bournemouth.ac.uk
Alan Grainger (Leeds): a.grainger@leeds.ac.uk
Junwoo Kim (Leeds): gyjk@leeds.ac.uk
Andy Lawrence (Royal Observatory, Edinburgh): al@roe.ac.uk
Gabriela Lopez Gonzalez (Leeds): sc14lga@leeds.ac.uk
Jim Lynch (Surrey): J.Lynch@surrey.ac.uk
Dave Morris (Royal Observatory, Edinburgh): dmr@roe.ac.uk
Keith Noddle (Royal Observatory, Edinburgh): ktn@roe.ac.uk
Oliver Phillips (Leeds): o.phillips@leeds.ac.uk
Shaun Quegan (Sheffield): s.quegan@sheffield.ac.uk
Tiina Sarkinen (Royal Botanic Garden, Edinburgh): t.sarkinen@rbge.ac.uk
Kevin Tansey (Leicester): kjt7@leicester.ac.uk
Guy Ziv (Leeds): g.ziv@leeds.ac.uk