1 Introduction

This report looks at *AstroTrop* project use casees and to determine the project requirements, and then evaluates where the concepts, components and services from the IVOA can be applied to meet some of the *AstroTrop* project requirements.

2 The IVOA Virtual Observatory

The *The International Virtual Observatory Alliance* (*IVOA*) was formed in June 2002 with a mission to

> "facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory."

The Virtual Observatory (VO) is the realization of the *IVOA* vision of an integrated and interoperating virtual observatory. The work of the IVOA focuses on the development of standards, providing a forum for members to debate and agree the technical standards that are needed to make the VO possible.

The operational VO itself is comprised of a global shared metadata registry, the Registry, and a number of individual data discovery and data access services deployed at each of the participating institutes. These components work together to present a uniform mechanism for discovering and accessing data, irrespective of where it is physically located.

The VO architecture and data discovery processes are very similar to the interconnected metadata collections approach described in "The new bioinformatics: integrating ecological data from the gene to the biosphere" (Jones et al. 2006).

> ".... a loosely structured collection of project-specific data sets accompanied by structured metadata about each of the data sets."

> "Each of the data sets is stored in a manner that is opaque to the data system in that the data themselves cannot be directly queried; rather, the structured metadata describing the data is queried in order to locate data sets of interest."

> "After data sets of interest are located, more detailed information .... can be extracted from the metadata and used to load, query, and manipulate individual data sets."

2.1 Example use case

A useful way to illustrate how the data discovery process works in the VO is to look at an example task such as selecting images covering a particular region of the sky, in a particular wavelength e.g. infrared, visible light, radio or xray.

2.1.1 Service discovery

The first step of the process is to identify the services that provide access to the type of data we are looking for by querying the *Registry*.

The *Registry* is comprised of a number of small local registry services, typically hosted at the participating institute level, working in cooperation with a set of higer level global registry services hosted by a few key institutes that aggregate the data from the smaller registries to create a global searchable index of metadata describing all of the services and datasets available in the VO.

When a new service is deployed, part of the deployment process involves registering the service with the local registry. The local registry is then responsible for collecting and storing the metadata that describes both the service itself and the datasets that it provides access to.

Once the metadata for a serviceor dataset has been registered in a local registry, it is automatically propagated up to the next level and replicated between the global registries.

This makes it possible to access the metadata for all of the services and datasets published in the VO by querying any one of the global registries.

The first step in fulfilling our example use case is to identify services that contain the type of data we are looking for, in this case images, by querying the *Registry* for services that support the *IVOA Simple Image Access* (*SIA*) capability.

In addition to the technical details of services and their capabilities the *Registry* also contains details about the content of datasets, including details of the wavelength(s) measured, e.g. infrared, visible, radio or xray.

This allows us to refine our query to search for *SIA* services that contain images in a specific waveband, e.g. optical, infrared or x-ray.

The *Registry* query returns a table of data, each row of which contains information about a *SIA* service that provides the type of data we are interested in - images in a particular wavelength.

The VO is itself an evolving system, building on the existing work to add additional levels of integration as new features are added to the IVOA specifications.

A recent addition to the list of *IVOA* standards is the *HEALPix Multi-Order Coverage Map* (*MOC*) which allows *Registry* services to perform coarse grained region matches.

This will enable us to further refine our *Registry* query to filter for *SIA* services that contained data in a particular region of the sky.

2.1.2 Data discovery

The next stage of the process is to query each of the *SIA* services in the list to discover details about the individual images available from that service.
A *SIA* service can handle queries that specifiy a particular wavelength and a particular region of the sky :

- `POS` The positional region (ra, dec)
- `BAND` The energy interval (wavelength)

Each *SIA* service returns a table of data, each row of which contains metadata about an individual image. The details of the fields in the image metadata are defined in the *Observation Data Model Core Components* (*ObsCore*) data model.

This demonstrates a core part of the *IVOA* architecture, interoperable services based on standard interfaces and data formats.

All of the *SIA* services will return a standard response, which makes it much easier to combine them to produce a global list of all the images available within the whole VO that match our search criteria.
The two key components of this are :

- A standard interface for the global *Registry* that uses a standard set of attributes to describe datasets and services
- A standard interface for local *SIA* data access services that uses a standard set of attributes to describe the available data products

The separation between the initial service discovery query at the global level followed by individual data discovery queries at the local level is very similar to the stages described in Jones et al. 2006 :

1. Querying the metadata to establish the location of suitable data
2. Querying the individual services to establish what the data is and how to access it

3 Tropical forest science

3.1 Carbon density comparison

We can compare the VO data discovery process for astronomy data with an example use case based on a recent study "Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites" (Mitchard et al. 2014), comparing remote sensing data from satellites with ground plot data collected in the field.

The study compares two sets of remote sensing data, from *NASA Jet Propulsion Laboratory* (*JPL*) "Benchmark map of forest carbon stocks in tropical regions across three continents" (Saatchi et al. 2011) [RS1] and the *Woods Hole*

Research Center (*WHRC*) “Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps” (Baccini et al. 2012) [RS2] with four sets of ground plot data from

- *Red Amazónica de Inventarios Forestales* (*RAINFOR*) (Peacock et al. 2007) (Malhi et al. 2009)
- *Amazon Tree Diversity Network* (*ATDN*)
- *Tropical Ecology Assessment and Monitoring* (*TEAM*)
- *Brazilian Program for Biodiversity Research* (*PPBio*) (Pezzini et al. 2012)

3.1.1 Remote sensing source data

It is not know what data discovery and data access methods were used to identify and access the primary remote sensing source data.

However, there are a number of data discovery tools available that enable researchers to search for remote sensing data products such as satellite images and radar scans.

A good examples of this type of tool are the *Earth Explorer* and *GloVis* tools provided by the *U.S. Geological Survey* (*USGS*)

> ”The USGS EarthExplorer ... provides users the ability to query, search, and order satellite images, aerial photographs, and cartographic products from several sources”

> ”In addition to data from the Landsat missions and a variety of other data providers, EE now provides access to MODIS land data products from the NASA Terra and Aqua missions, and ASTER level-1B data products over the U.S. and Territories from the NASA ASTER mission”

> ”The USGS Global Visualization Viewer (GloVis) is an online search and order tool for selected satellite data. The viewer allows access to all available browse images from the Landsat 7 ETM+, Landsat 4/5 TM, Landsat 1-5 MSS, EO-1 ALI, EO-1 Hyperion, MRLC, and Tri-Decadal data sets, as well as Aster TIR, Aster VNIR and MODIS browse images from the DAAC inventory”

The *USGS* also provides large area composited mosaics generated from *Landsat* data via the *WELD* project.

> ”The WELD data products are processed so users do not need to apply the equations, spectral calibration coefficients, and solar information, needed to convert Landsat digital numbers to reflectance and brightness temperature. They are defined in the same coordinate system and align precisely, making them simple to use for

> multi-temporal applications. The products provide consistent data that can be used to derive higher-level land cover as well as geophysical and biophysical products for assessment of surface dynamics and to study Earth system functioning"

The *USGS* also maintains a *Long Term Archive* (*LTA*) of historical remote sensing data.

> "The U.S. Geological Survey's (USGS) Long Term Archive (LTA) at the National Center for Earth Resource Observations and Science (EROS) in Sioux Falls, SD is one of the largest civilian remote sensing data archives"

> "Time series images are a valuable resource for scientists, disaster managers, engineers, educators, and the general public. USGS EROS has archived, managed, and preserved land remote sensing data for more than 35 years and is a leader in preserving land remote sensing imagery"

However, all of these interfaces are based around human interaction. There are no machine readable data discovery services for this type of remote sensing data.

3.1.2 Carbon density maps

A detailed description of the [RS1] dataset produced by *NASA Jet Propulsion Laboratory* is available in the authors paper (Saatchi et al. 2011).

The paper, along with the additional supporting information available on the *PNAS* website, describes the main upstream data sources and the methods applied.

However, details of the data sources, the instruments, target areas and date ranges the data covers are not available in a machine readable format.

> "Ground data used to train the biomass prediction model were derived from various sources including published literature and national forest inventories collected by the authors and their colleagues."

The carbon density dataset itself is available as *GTIF* files, with associated *World file* metadata, for download from the *NASA JPL carbon dataset* site.

A detailed description of the [RS2] dataset produced by the *Woods Hole Research Center* is available in the authors paper (Baccini et al. 2012).

The paper, along with the additional supporting information available from the *Nature* website, describes the upstream data sources and the methods applied. However, details of the data sources, the instruments, target areas and date ranges the data covers are not available in a machine readable format.

The carbon density dataset itself is available by request from the *WHRC carbon dataset* website. Access to the data requires filling in a simple web form declaring who you are and what you want to use the data for. On submitting the web-form, an automated email reply is generated containing a URL to a *ZIP* file on the WHRC website.

The *ZIP* file contains the data as *GTIF* files, with associated *World file* metadata.

3.1.3 Ground plot data

The four sets of ground plot data from *RAINFOR*, *ATDN*, *TEAM* and *PPBio* were combined together in the *ForestPlots.Net* database.

Details of the design and capabilities of the *ForestPlots.Net* system is presented in "ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data" (Lopez-Gonzalez et al. 2011).

> "The ForestPlots.net web application was designed primarily as a repository for long-term intact tropical forest inventory plots, where trees within an area are individually identified, measured and tracked through time"

Of the three sets of ground plot data, the data from *RAINFOR* and *ATDN* were already available in the *ForestPlots.Net* database.

The plot data from the *TEAM* and *PPBio* projects were downloaded and imported into the *ForestPlots.Net* database manually.

A permanent archive of the combiuned field plot data is stored in the *ForestPlots.Net* database as a publically available dataset[1] and is available in the supporting information for the paper.

3.1.4 *AGB* data

The *AGB* data for the forest plots were calculated using a *SQL* query provided by the *ForestPlots.Net* system which implements the tropical forest model described in "Tree allometry and improved estimation of carbon stocks and balance in tropical forests" (Chave et al. 2005). The results of the *AGB* calculation for each forest plot are included in the combined field plot dataset stored in the *ForestPlots.Net* database.

The paper refers to a number of maps derived from the field plot data and other sources which were generated as part of the analysis.

- *Kriged* map of mean wood density (ρ)
- Ratio of diameter (D) to tree height (H) Feldpausch et al. 2012
- *Kriged* map of basal area
- *Kriged* map of *AGB* using D and species-specific ρ, and a regional height model ($K_{DH\rho}$)

[1] `http://dx.doi.org/10.5521/FORESTPLOTS.NET/2014_1`

- *Kriged* map of *AGB* using D and species-specific ρ, but a pan Amazonian height model ($K_{D\rho}$)
- *Kriged* map of *AGB* using D, regional height models and ρ, but with ρ fixed at 0.63 (K_{DH})
- *Kriged* map of *AGB* using D, pan-Amazonian height model, and ρ fixed at 0.63 (K_D)
- *AGB* map from [RS1] (Saatchi et al. 2011)
- *AGB* map from [RS2] (Baccini et al. 2012)
- Difference between [RS1] and $K_{DH\rho}$
- Difference between [RS2] and $K_{DH\rho}$
- Difference between [RS1] and [RS2]

These derived datasets and maps are not available in the supporting information for the paper.

The *AGB* data derived from two remote-sensing-derived maps, Saatchi et al. 2011 [RS1] and Baccini et al. 2012 [RS2] are not available in the supporting information for the paper.

4 Data discovery requirements

Based on the *AstroTrop* use cases we have studied so far it is clear that data discovery forms a significant part of the requirements for *AstroTrop*.

In the many of the use cases a significant part of the source material for the use case has come from outside the *AstroTrop* community.

For example, both the [RS1] (Saatchi et al. 2011) and [RS2] (Baccini et al. 2012) datasets used in the Mitchard et al. 2014 use case came from external data sources, *NASA Jet Propulsion Laboratory* and *Woods Hole Research Center* respectively.

In the long term, the ideal solution would be to encourage these external data providers to join the *AstroTrop* community and publish and curate their own data using the *AstroTrop* system.

In the short term, in order to make this type of external data available as part of the *AstroTrop* data discovery process, it will necessary for a member of the *AstroTrop* community to register and curate the metadata about the external data in the *AstroTrop* system.

A number of the *AstroTrop* use cases also require data created or provided by members of the *AstroTrop* community.

This means that the *AstroTrop* needs to provide support for individual members of the *AstroTrop* community to publish indivudual datasets within the *AstroTrop* system. Enabling other members of the *AstroTrop* community to discover and use the data as source material for their own research. Promoting the sharing and re-use of results within the *AstroTrop* community.

In both cases, the requirements for the data discovery process are that the users is able to specify an area of interest and the type of data they are interested in and then gradually narrows the search criteria in response to the data discovery results until they find the data they are interested in.

Based on this outline we can begin to evaluate how well the *IVOA* and *AstroGrid* software meets the *AstroTrop* requirements and compare this with equivalent *GIS* software currently available.

At first glance, the *IVOA* and *AstroTrop* data discovery processes are very similar. Suggesting that the *IVOA* and *AstroGrid* software should be a good fit for the *AstroTrop* requirements.

However, there are a number of issues that may mean that the *IVOA* and *AstroGrid* software are not the most best solution for meeting the *AstroTrop* requirements.

One issue is that a significant part of the *IVOA* metadata structure relies on a number of domain specific astronomy terms and concepts, making it difficult to apply the *IVOA* metadata structure to data from a different domain.

Although it would be possible to remove the domain specific terms and concepts from the *IVOA* data model and replace them with a set of terms that were more suited to the *AstroTrop* domain. Doing this piece at a time, gradually evolving a new metadata data model for the *AstroTrop* project would be a non-trivial undertaking involving a significant comitment of time and resources. It is worth bearing in mind that the *IVOA ObsCore* data model that forms the basis of a lot of the *IVOA* data discovery process is the result of 10 years work by the *IVOA* working groups to define a common data model for astronomy observations.

With this in mind, it would probably be more practical to base the *AstroTrop* metadata data model on a existing data model(s) that has been developed for domains similar to *AstroTrop*.

Two examples of this are the *World file GIS metadata* and *Ecological Metadata Language* (*EML*) metadata formats.

The *World file GIS metadata* format provides a simple way of annotating an existing map or raster image with GIS location metadata.

The *World file* format consists of a plain text file format containing details of the location, scale and rotation of a map or raster image.

Both of the remote sensing datasets [RS1] (Saatchi et al. 2011) and [RS2] (Baccini et al. 2012) provide *World file* metadata using the *example.tfw* convention to associate the metadata with the *GTIF* maps.

Ecological Metadata Language (*EML*) is a metadata specification for describing ecological datasets, based on work done by the Ecological Society of America and associated efforts "Nongeospatial metadata for the ecological sciences" (Michener et al. 1997).

EML is implemented as a series of *XML* document types that can be used to describe different aspects of an ecological dataset.

A second issue with the *IVOA* and *AstroGrid* metadata *Registry* and data discovery tools concerns the allocation of roles and responsibilities for managing the metadata within the *Registry*, and the way these reflect the structure of the *IVOA* and its members involved in the development process.

Historically, the most active contributors to the development of the *IVOA* standards and the *AstroGrid* software have been primary data providers within the astronomy community.

Many of these represent large scale data providers responsible for publishing and curating the primary science archive for a telescope survey or satelite mission.

In the *AstroTrop* domain these are equivalent to the upstream data providers who publish and curate the original satelite remote sensing data, such as the *Landsat* data archive or the *LTA* provided by the *USGS*.

This has influenced the way that the *IVOA* and *AstroGrid* software has been developed. In particular, the initial emphasis on providing the tools and services for publishing large datasets, meant that the curation of the dataset metadata was seen as a system administrator role, responsible for curating all of the metadata published by that institute.

As a result, many of the current user interface tools for managing and curating data sets are designed around a single system administrator managing the metadata for an entire service, rather that individual researchers managing the metadata for their own data.

In contrast, in the *AstroTrop* use cases a significan portion of the data in the system will be provided by and curated by individual researchers or small research groups. For example, the results and supplementary data for the Mitchard et al. 2014 paper would be published and curated by the members of the research team themselves.

Neither of these issues are 'show stoppers', it should be possible to replace the *IVOA ObsCore* with a a new datamodel designed for *AstroTrop*, and it should be possible to develop a new user interface and permission infrastructure to enable individual users to publish and curate their own data.

However, it would be useful at this point to compare the data discovery capabilities and features provided by the *IVOA* and *AstroGrid* software with the data discovery capabilities of some of the *GIS* based systems designed to handle geographical and ecological data.

4.0.5 Global Index of Vegetation-Plot Databases

The *Global Index of Vegetation-Plot Databases* (*GIVD*) system is a complex registry of metadata describing databases of vegetation plot data from around the world.

The *GIVD* system contains records for *... 209 databases with 3,148,605 vegetation plots*, including three of the datasets used in our use case.

- *ForestPlots.Net* `[GIVD:00-00-001]`[2]

[2] `http://www.givd.info/ID/00-00-001`

- *PPBio* `[GIVD:SA-BR-001]`[3]
- *TEAM* `[GIVD:00-00-002]`[4].

In "The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science" (Dengler et al. 2011) the *GIVD* project team describe the system architecture and outline plans to aggregate different types of data from external sources.

> "Our longer-term vision is to develop GIVD in ways similar to Metacat (Jones et al. 2006), so that, ultimately, users who query GIVD will not only receive information on which databases contain data suitable for the intended analyses, but they will also discover other data from distributed databases, with GIVD acting as the central node."

This is broadly similar to the VO architecture of distributed datasets the interconnected metadata collections approach described in "The new bioinformatics: integrating ecological data from the gene to the biosphere" (Jones et al. 2006).

However, the emphasis is on providing a human interactive search facility with the *GIVD* system acting as the central node. The plans do not include providing a machine readable interface, allowing the *GIVD* system itself to be used as a component in in a larger distributed system.

4.0.6 *PPBio* Information System

In "The Brazilian Program for Biodiversity Research (PPBio) Information System" (Pezzini et al. 2012) the *PPBio* team describe the role of the data manager and the metadata collection processes developed as part of the *PPBio* Information System.

They also describe the transition from an initial flat file data storage system, to a new system based on *Metacat*.

> "To facilitate data searches, all the metadata were converted to XML, and the PPBio has installed a METACAT server to integrate with the Knowledge Network for Biocomplexity (KNB), a network which aims to assist ecological and environmental research."

This indicates a move towards using open standards for the metadata and the service interfaces, enabling the *PPBio* Information System to become part of a larger distrbuted system.

[3] `http://www.givd.info/ID/SA-BR-001`
[4] `http://www.givd.info/ID/00-00-002`

4.0.7 *Knowledge Network for Biocomplexity*

The *Knowledge Network for Biocomplexity KNB* is a data repository

> "intended to facilitate ecological and environmental research"

by enabling researchers to

> "share, discover, access and interpret complex ecological data"

The *KNB* system uses the *Metacat* software to store and query the *Ecological Metadata Language* (*EML*) metadata for each of the datasets in the repository.

In some cases the *KNB* system stores both the metadata and actual data itself, e.g. *Tree crown allometries, Piedmont and Southern Appalachians 2001-2004*[5].

In other cases the *KNB* system only stores the metadata, referring to data that is stored elsewhere, e.g. *Tree crown allometries, Piedmont and Southern Appalachians 2001-2004*[6].

4.0.8 *Metacat*

Metacat is an open source data management tool that provides a repository for managing both data and metadata in a single system.

> Metacat is a repository for data and metadata (documentation about data) that helps scientists find, understand and effectively use data sets they manage or that have been created by others.

Metacat is capable of handling a variety of different metadata formats, including *Ecological Metadata Language* (*EML*)

4.0.9 DataONE

The Metacat project is itself part of the Data Observation Network for Earth (DataONE) project, a collaboration spconsored by the U.S. National Science Foundation to build an infrastructire from distributed webservices that provides open, persistent, robust, and secure access to Earth observational data.

> The DataONE project is a collaboration among scientists, technologists, librarians, and social scientists to build a robust, interoperable, and sustainable system for preserving and accessing Earth observational data at national and global scales. Supported by the U.S. National Science Foundation, DataONE partners focus on technological, financial, and organizational sustainability approaches to building a distributed network of data repositories that are fully interoperable, even when those repositories use divergent underlying software and support different data and metadata content standards.

[5] `https://knb.ecoinformatics.org/#view/doi:10.5063/AA/mdietze.3.2`

[6] `https://knb.ecoinformatics.org/#view/doi:10.5063/AA/mdietze.3.2`

The DataONE arcitecture is based on a set of top level *Coordinating Nodes* and *Member Nodes* located at each participating institute or organisation

Coordinating Nodes provide a replicated catalog of Member Node holdings, enabling scientists to discover data wherever they reside, and data repositories to make their data and services available to the international community.

The individual *Member Nodes* at each institute enable them to make their data available to the rest of the DataONE network via a standard webservice interface.

Again, this two layers of data discovery and data access is similar the virtual observatory architecture.

4.0.10 ***FGDC Geospatial Platform***

http://ckan.org/features/

> "Many organisations already have their data in repositories with well-defined process and procedures for publishing and managing data. In this case the data can be simply pulled regularly into CKAN from the existing repositories. To facilitate this model we've developed a sophisticated and customisable "harvesting" mechanism which can fetch and import records from many different repository sources, including:

- Geospatial CSW Servers (see geospatial for more information)
- Existing web catalogues
- Simple HTML index pages or Web Accessible Folders
- ArcGIS, Geoportal Servers and Z39.50 databases
- Other CKAN instances

> This functionality is used on data.gov to get data in from hundreds of their agencies, on data.gov.uk to implement a Discovery Metadata Service used to fulfill the UK's obligations under the EU INSPIRE directive."

> Because the harvesting functionality can be used to pull in metadata from other CKAN instances, it can also be used to create a federated network of CKAN nodes which share data between each other. This is useful if, for example, a national portal wanted to aggregate information from local government CKAN instances, or if a topic-specific CKAN instance was created which aggregated a subset of datasets from other CKAN sources. CKAN follows the DCAT standard for data catalogue metadata, so data can also be federated from other non-CKAN catalogues."

GBIF

Uses EML Uses Hadopop internally Exposes RESTful API allowing external access. Automated injestion and processing pipeline. Apache SOLR for federated search.

AstroTrop requirements

Data discovery for primary data (NASA, ESO, WELD). Data discovery for external data (JPL, WHR).

Data discovery for community data (ForestPlots, PPBio).

Data discovery for external data (GIVD, KNB, GBIF).

Data discovery for community data (Mitchard, Tansey).

Key requirement is GIS location data, region() etc.

AstroGrid services

Registry

XML registry - fixed metadata format based on IVOA VOResource TAP registry - fixed metadata format based on IVOA VOResource MOC support not implemented

SIAP

GIS not implemented, AG services only provide simple box, not cone.

TAP

Updated service is work in progress Can be adapted to meet AT requirements GIS support not implemented on SQLServer. OpenGIS equivalent.

References

AGB. Above Ground Biomass. URL: `http://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_04_Ch4_Forest_Land.pdf`.

ATDN. Amazon Tree Diversity Network. URL: `http://web.science.uu.nl/Amazon/ATDN/`.

AstroGrid. URL: `http://www.astrogrid.org/`.

AstroTrop. URL: `http://www.astrotrop.org/`.

Baccini, A. et al. (2012). "Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps". In: *Nature Climate Change* 2 (3), 182–185. DOI: `10.1038/nclimate1354`. URL: `http://www.nature.com/nclimate/journal/v2/n3/full/nclimate1354.html`.

PPBio. Brazilian Program for Biodiversity Research. URL: `http://www.teamnetwork.org/`.

Chave, J. et al. (2005). "Tree allometry and improved estimation of carbon stocks and balance in tropical forests". In: *Oecologia* 145 (1), pp. 87–99. DOI: `10.1007/s00442-005-0100-x`. URL: `http://link.springer.com/article/10.1007/s00442-005-0100-x`.

Dengler, Jürgen et al. (2011). "The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science". In: *Journal of Vegetation Science* 22 (4), pp. 582–597. DOI: `10.1111/j.1654-1103.2011.01265.x`. URL: `http://onlinelibrary.wiley.com/doi/10.1111/j.1654-1103.2011.01265.x/abstract`.

Dietze, Michael, Michael Wolosin, and James Clark (2004). *Tree crown allometries, Piedmont and Southern Appalachians 2001-2004*. URL: https://knb.ecoinformatics.org/#view/doi:10.5063/AA/mdietze.3.2.

Earth Explorer. URL: https://lta.cr.usgs.gov/earth_explorer.

EML. Ecological Metadata Language. URL: https://en.wikipedia.org/wiki/Ecological_Metadata_Language.

XML. Extensible Markup Language. URL: http://www.w3.org/XML/.

Feldpausch, T. R. et al. (2012). "Tree height integrated into pantropical forest biomass estimates". In: *Biogeosciences* 9 (8), 3381–3403. DOI: 10.5194/bg-9-3381-2012. URL: http://www.biogeosciences.net/9/3381/2012/bg-9-3381-2012.html.

FGDC Geospatial Platform. URL: http://www.geoplatform.gov/.

ForestPlots.Net. URL: http://www.forestplots.net/.

Kriged. Gaussian process regression. URL: https://en.wikipedia.org/wiki/Kriging.

GIS. geographic information system. URL: http://en.wikipedia.org/wiki/Geographic_information_system.

GTIF. GeoTIFF. URL: http://trac.osgeo.org/geotiff/.

GIVD. Global Index of Vegetation-Plot Databases. URL: http://www.givd.info/.

GloVis. Global Visualization Viewer. URL: https://lta.cr.usgs.gov/glovis.

MOC. HEALPix Multi-Order Coverage Map. URL: http://www.ivoa.net/documents/MOC/.

Registry. IVOA Registry. URL: http://www.ivoa.net/documents/RegistryInterface/.

SIA. IVOA Simple Image Access. URL: http://www.ivoa.net/documents/SIA/.

Jones, Matthew B. et al. (2006). "The new bioinformatics: integrating ecological data from the gene to the biosphere". In: *Annual Review of Ecology, Evolution, and Systematics* 37, pp. 519–544. DOI: 10.1146/annurev.ecolsys.37.091305.110031. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.37.091305.110031.

KNB. Knowledge Network for Biocomplexity. URL: https://knb.ecoinformatics.org/.

Landsat. Landsat Program. URL: http://landsat.gsfc.nasa.gov/.

LTA. Long Term Archive. URL: https://lta.cr.usgs.gov/about.

Lopez-Gonzalez, Gabriela et al. (2011). "ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data". In: *Journal of Vegetation Science* 22 (4), pp. 610–613. DOI: 10.1111/j.1654-1103.2011.01312.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1654-1103.2011.01312.x/abstract.

Malhi, Y. et al. (2009). "An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR)". In: *Journal of Vegetation Science* 13 (3), pp. 439–450. DOI: 10.1111/j.1654-1103.2002.tb02068.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1654-1103.2002.tb02068.x/abstract.

Metacat. Metacat. URL: https://knb.ecoinformatics.org/knb/docs/intro.html.
Michener, William K. et al. (1997). "Nongeospatial metadata for the ecological sciences". In: *Ecological Applications* 7 (1), pp. 330–342. DOI: 10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2. URL: http://www.esajournals.org/doi/abs/10.1890/1051-0761%281997%29007%5B0330%3ANMFTES%5D2.0.CO%3B2.
Mitchard, Edward T. A. et al. (2014). "Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites". In: *Global Ecology and Biogeography* 23 (8), pp. 935–946. DOI: 10.1111/geb.12168. URL: http://onlinelibrary.wiley.com/doi/10.1111/geb.12168/abstract.
JPL. NASA Jet Propulsion Laboratory. URL: http://carbon.jpl.nasa.gov/.
NASA JPL carbon dataset. URL: ftp://www-radar.jpl.nasa.gov/projects/carbon/datasets/.
Nature. URL: http://www.nature.com/.
ObsCore. Observation Data Model Core Components. URL: http://www.ivoa.net/documents/ObsCore/.
Peacock, J et al. (2007). "The RAINFOR database: monitoring forest biomass and dynamics". In: *Journal of Vegetation Science* 18 (4), 535–542. DOI: 10.1111/j.1654-1103.2007.tb02568.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1654-1103.2007.tb02568.x/abstract.
Pezzini, Flávia Fonseca et al. (2012). "The Brazilian Program for Biodiversity Research (PPBio) Information System". In: *Biodiversity & Ecology* 4 (24), 265–274. DOI: 10.7809/b-e.00083. URL: http://www.biodiversity-plants.de/biodivers_ecol/article_meta.php?DOI=10.7809/b-e.00083.
PNAS. Proceedings of the National Academy of Sciences of the United States of America. URL: http://www.pnas.org/.
RAINFOR. Red Amazónica de Inventarios Forestales. URL: http://www.rainfor.org/.
Saatchi, Sassan S. et al. (2011). "Benchmark map of forest carbon stocks in tropical regions across three continents". In: *Proceedings of the National Academy of Sciences of the United States of America* 108 (24), 9899–9904. DOI: 10.1073/pnas.1019576108. URL: http://www.pnas.org/content/108/24/9899.
SQL. Structured Query Language. URL: https://en.wikipedia.org/wiki/SQL.
IVOA. The International Virtual Observatory Alliance. URL: http://www.ivoa.net/.
TEAM. Tropical Ecology Assessment and Monitoring. URL: http://www.teamnetwork.org/.
USGS. U.S. Geological Survey. URL: http://www.usgs.gov/.
WELD. Web-enabled Landsat data. URL: http://landsat.usgs.gov/WELD.php.
WHRC carbon dataset. URL: http://www.whrc.org/mapping/pantropical/carbon_dataset.html.

WHRC. Woods Hole Research Center. URL: http://www.whrc.org/.
World file. World file GIS metadata. URL: https://en.wikipedia.org/wiki/World_file.
ZIP. ZIP archive. URL: https://en.wikipedia.org/wiki/Zip_(file_format).