UDC: 574.1

Review paper

Overview of Free Open Source Global Forest Species Data for Biogeographic Modeling

Teo Beker *, Saša Orlović, Dejan B. Stojanović

University of Novi Sad, Institute of Lowland Forestry and Environment, Novi Sad, Serbia

* Corresponding author: Teo Beker; E-mail: teo.beker@uns.ac.rs

Abstract: In forestry a lot of time and effort is spent in forest monitoring and conservation, and until recently there has not been available a successful, economical and easy way to automate this. This implies great importance in sharing the collected data and using it efficiently. Today it is becoming more common to publish the collected data sets from previous studies as the free open source data. Thanks to that data does not have to be collected twice for the same region, and researchers can focus on using the data and developing and testing different models without the need to go onsite, set the experiments and collect the data for years or even longer, which was the case before. This also gives the option of making informed decisions on new regions and testing models before stepping into the multiple year projects or experiments. This study aims to give overview of the free open source data covering global or continental level for use in forestry. Two types of data are described: the primary data and secondary data (derived maps). The biggest biodiversity and biologic collection databases have been selected and described. The brief analysis of the data sources is given and primary sources for future research have been pointed out.

Keywords: biodiversity, biogeographic modeling, open source, forests, spatial data.

1. Introduction

Forest monitoring and data collection is extremely slow and time-consuming process. Usually it includes specialists going on site with GNSS and measuring individual trees and marking the location. There are newer techniques which can be used for less precise measurements, like Lidar and Remote Sensing using drones or satellites. Lidar (Akay et al. 2009) or remotely sensed (Jovanović et al. 2015; Tang and Shao, 2015) data should be preprocessed before the useful data is extracted. Data received in any of these methods can be used in combination with others, and for that to be possible, some standards must be followed. There are many local approaches with a few attempts to standardize it on global scale. Usually in Serbia the forest inventories are regulated on the country level, or per owner. There is a lot of variations on the process or the type of collected data, depending on the question being asked (Feinsinger, 2001). Food and Agriculture Organization of United Nations has tried to systematize the process (Tomppo and Andersson, 2008). They have created guidelines for sampling the forest data in different conditions and contexts. After the data is collected, often it is stored in excel sheets, or much more rarely in databases or GIS systems. These are often kept on personal computers at inaccessible locations. It is slowly catching on to publish the data on online platforms for wider accessibility. Still even when published, often this data is kept on local: scientific, governmental or private lesser-known web platforms, less often unified in larger, well known global ecosystem.

For users to know what they are looking for metadata is of extreme significance. Metadata gives information how the data was acquired, when, where etc. This is very important information for the data to be used properly. If the data is meant to be published on global ecosystem the standardization plays even larger role. Ability to search data efficiently and to find the needed set among billions of published data sets depends on the use and standardization of metadata. The standardization of such metadata also proved to be a problem. For biodiversity purposes, Darwin Core (Wieczorek et al. 2012) system has been created and has started gaining traction. It is based on Dublin core (Weibel, 1997) metadata specification which is simplified cataloging specification for digital content. There are also other general metadata cataloging specifications still in use like MARC (Avram, 2003) and AARC2 (Hjerppe and Olander, 1989).

The application uses of collected data can be wide, it can range from decision making on primary data to modeling and generating secondary data. Biogeographic modeling (Guisan et al. 2017) has developed for a long time now, and many useful models and theories have matured (Jørgensen et al. 2009). It is common for all those models that they need the data for training and calibrating them, also many such models benefit from larger amounts of data and wider area cover, optimally whole range of habitats the species covers. In past this this was almost impossible, once collected data until recently did not see much reuse. Today, there is huge improvement in this field, people who publish open access data are being recognized and cited for their data sources, there are platforms which support publishing process and integrate different data sets, and the culture of sharing is developing. In this and other ways open access data and open science concepts are being supported and stimulated.

With the rise of internet, scientists saw the potential and around 2000s in many parts of the world the process of structuring species data and creating a platform for collecting such data has started (Canhos et al. 2004). As time passed some platforms proved more successful than others, other merged and created larger integrated databases (Catapano et al. 2001; Shao et al. 2007) and today biological scientists have a large choice of platforms or data sources to choose from. Within these platforms the species data can be found in raw state, in a form of points where samples of certain species have been marked and may include valuable additional information related to for example monitoring process, precision and accuracy of the given data set. Also data can be found in a form of maps depicting species distribution, distribution of species density, habitat, species health or already modeled species potential niches. In this paper both types: primary (raw) and secondary (processed data) are analyzed.

The goal of this paper is to summarize the sources of the biological species information, to briefly compare it on basic criteria like number of occurrences, data sets and species contained on platforms, and the way they have been organized and developed. By doing so motivate users to use more the existing biodiversity databases infrastructure and allow further specialization between data collection and modeling professions. The paper covers two parts, one presenting the biodiversity platforms with the the primary species data and other presenting the studies using the primary data to create useful derivatives and maps (secondary data).

2. Primary (Raw) Biological species databases

In this section the platforms which collected spatial biodiversity information on global level are presented. Only the main characteristics of platforms where forestry data, among other species data, can be found are described.

2.1. GBIF

Global Biodiversity Information Facility (GBIF) is the biggest and one of the most influential global source of biodiversity data. It is an international network and research infrastructure funded by worlds governments. Its idea was formed in 1999 and the project was officially established in 2001. It is said that GBIF is the largest online provider of biodiversity distribution records. It was developed to give access to data for everyone interested in living beings on Earth. Participating countries provide data which is prepared with common standards. It has many sources, from museum specimens two century old, to geo-tagged smartphone photos. It combines all the data using Darwin Core Standards, which is the basis of indexing of records.

At the time of writing GBIF database contains almost 1.3 billion occurrence records, more than 46 thousands data sets and 1,473 publishing institutions (GBIF, 2019). It can be accessed at https://www.gbif.org/.

2.1.1. iNaturalist

iNaturalist is the platform developed by joint initiative of California Academy of Sciences and National Geographic society. It is an online social network for sharing biodiversity information. It uses crowd-sourced species identification system and an organism recording tool. That means that it is possible to get the data for analysis, publish data, or help other users with identification of the species. It consists of web and mobile platform to be accessible at any place and time. The goal of the project is to connect people to nature and produce scientifically valuable biodiversity data. At the time of writing it had 28 Million observations and 236 thousand species and almost 2 million users. It can be accessed at https:// www.inaturalist.org/. The data from iNaturalist are indexed and included in GBIF.

2.2. BIEN

More than 50 people from multiple universities and institutions across United States of America united in 2008 to create BIEN - Botanical Information and Ecology Network, a web platform to bring together disparate networks of botanical researchers. The goals of this structure are:

- To connect the leading collectors of botanical survey and inventory data, information and computer scientists, ecologists, biologists and other scientists interested in plants research,
- To merge existing and collect new global botanical observation data,
- To establish the informatics infrastructure for the community to stimulate the discovery, study and preservation of botanical diversity.

It is estimated based on Index Herbariorum that world museums collected in last 400 years and hold more than 380 million plant specimens. The specimens are steadily being digitized, most of them still not available to be accessed digitally. This fact alone implies the importance and significance of biodiversity platforms.

The collection data, vegetation plot records and plant traits, phylogeny as well as other potentially useful information is combined from many data sources, checked and presented on BIEN platform (Enquist et al. 2016).

Since version 3, BIEN has been rebuilt from the ground up to use Darwin Core standards exchange schema. At the time of writing BIEN is in version 4.1, has about 200 million observations of around 60 million specimens and 485 thousands species. It can be accessed at http://bien.nceas.ucsb.edu/bien/.

2.3. RainBio

The RainBio project was started in 2014 to create a high quality observations of vascular plants in sub-Saharan tropical region, including Gulf of Guinea islands, Cape Verde and Zanzibar archipelagos. The project ended in December 2016.

The RainBio database consists of 610,117 geolocated occurrences for 25,356 species of vascular plants and 29,659 taxa (including subspecies and varieties), 3,158 genera and 273 families, observed from 1782 to 2015. It can be accessed and downloaded at http://rainbio.cesab.org/. It follows the Darwin Core standard. The database is the compilation of thirteen data sets from world institutions, which have had removed the duplicates and been checked for the quality (Dauby et al. 2016).

2.4. DryFlor

Latin American Seasonally Dry Tropical Forest Floristic Network (DryFlor) is a continental scientific community network created to carry out biogeographic analysis for research and conservation of neotropical dry forests. Seasonally dry forests are found throughout Latin America, from Mexico to south-western Brazil and northern Argentina. They are characterized by a long period of drought, complete canopy and lack of grassy ground vegetation. Dry forests are highly threatened, currently they cover less than 10% of their original extent (K Banda-R DRYFLOR, 2016).

Dry or network has collected data from 1,600 floristic surveys covering 4,660 species of woody plants from the entire neotropical region, combined them and published as a database on their website. The database is focused on plants which grow more than 3m in height excluding lianas and climbers. It follows the nomenclature of the Angiosperm Phylogeny Group classification (APG III) for families. At the time of writing database contained 208,324 observations. It can be accessed at http://rainbio.cesab.org/.

2.5. Atlas of Living Australia

Atlas of Living Australia (ALA) is Australia's national biodiversity database. It is open for access for everyone. The project was initiated in 2006 (Belbin and Williams, 2016).

The web portal was based on GBIF data portal with certain improvements for ease of search, use and access of data. As it is based on GBIF, it uses Darwin Core standard as well. Many tools have been developed, and metadata repository created. Data integration of 10 scientific institutions was done by 2010, and later kept updating (Booth et al. 2012).

At the time of writing it has 86 million records of 124 thousands species in more than 11,500 data sets. It can be accessed at https://www.ala.org.au/.

2.6. Encyclopedia of Life

Encyclopedia of Life is free online platform meant to collect and distribute the knowledge and data on life-forms on Earth, and to increase awareness and understanding of living beings. It is hosted by Smithsonian National Museum of Natural History.

The idea of Encyclopedia of Life was born in 2003, with development of technologies for faster, and more systematic creation and collection of data on biodiversity and particular species in one place and for dissemination and presentation of the knowledge across the world (Wilson, 2003). For the project to take traction it took five more years, and in 2008 the first version of the Encyclopedia of Life went live. Currently it is in its third version which has structural improvements implemented, new functionalities, as forum, open data portal and has also been designed for mobile devices first. It is also based on Darwin Core standards which helps easily integrate it with other platforms and data.

In 2014 it had 3.5 million distinct pages for taxa and provides content for 1.3 million of those pages (Parr et al. 2014). Since then it has kept growing. It contains some spatial data, which unfortunately still is not abundant, but it makes up for that in different data sets, some of which are plant trait databases for main regions, species lists, Environment EOLs, Global Biotic Interactions, Summarized records, etc. It can be accessed on https://eol.org/.

3. Secondary (derived) data

Secondary or derived data is often modeled using some of the mentioned primary data sources of biodiversity data or remote sensing techniques. It is often in raster format or sometimes in vector polygon formats. Here there are mentioned three sources of secondary biodiversity data and the data they offer.

3.1. Copernicus

Copernicus Monitoring service offers users to download pan-European Corine land classification, Forest Type maps, Tree Cover Density maps and Dominant Leaf Type maps among other cover layers. Corine offers status maps from 1990, 2000, 2006, 2012 and 2018, while other offer maps in 2012 and 2015.

3.1.1. Corine

The Corine Land Cover (CLC) inventory was initiated in 1985. It classifies the land cover into 44 classes, among which broad-leaved, coniferous and mixed forests. CLC uses a Minimum Mapping Unit (MMU) of 25 hectares (ha) for areal phenomena and a minimum width of 100 m for linear phenomena. Also available are the change layers which are supplied in higher resolution, with MMU of 5 hectares. It is produced using different Satellite data, and in some countries using land survey data and national GIS (Bossard et al. 2000) .

Time consistency has been improving from the start of the programme, and it amounted to maps being made from 12 years of satellite imagery in the beginning, and in last instance of CLC 2018 it amounted to about a year. Geometric accuracy of the satellite data also shows the similar improvement, in CLC 1990 it was less than 50m on Landsat 5, and on CLC 2018 it is less than 10m on Sentinel 2 and Landsat 8 which are used. Aimed thematic accuracy is more than 85%, and since CLC 2000 it should be consistently achieved. Production time has been significantly lowered with better software and higher computing power, from 10 years in the beginning to 1.5 years for CLC 2018. Also number of countries included in the project has been increasing and for last two mappings amounts to 39 countries.

3.1.2. High Resolution Layers

The High Resolution Layers (HRL) Forests includes 3 types of products which are available for 2012 and 2015 reference years:

- Tree cover density (TCD) (level of tree cover density in a range from 0-100%),
- Dominant leaf type (DLT) (broadleaved or coniferous majority),
- A Forest type product (FTY).

The forest type product allows to get as close as possible to the Food and Agriculture Organization of United Nations (FAO) forest definition. In its original (20m) resolution it consists of two products:

- A dominant leaf type product that has a MMU of 0.5 ha and 10% tree cover density threshold applied
- A support layer that maps, based on the dominant leaf type product, trees under agricultural use and in urban context (derived from CLC and imperviousness 2009 data).
- For the final 100m product, trees under agricultural use and urban context from the support layer are removed (Ramminger et al. 2017).

3.1.3. Copernicus Global Land Service

The Copernicus Global Land Service (CGLS) is a component of the Land Monitoring Core Service (LMCS) of Copernicus. The Global Land Service systematically produces a series of bio-geophysical products on the status and evolution of the land surface, at global scale and at 300m to 1km spatial resolution, at daily, weekly or monthly intervals. It offers Burnt Area maps, NDVI (normalized difference vegetation index), Dry Matter Productivity, Soil Water Index, FAPAR (Fraction of Absorbed Photosynthetically Active Radiation), Surface Soil Moisture, FCOVER (Fraction of green Vegetation Cover), VCI (Vegetation Condition Index), Leaf Area Index, VPI (Vegetation Productivity Index), Land Cover etc.

3.2. Global Forest Watch

Global Forest Watch (GFW) is an international data and mapping network whose goal is to provide accurate information about world forests and by that include transparency and accountability in decision making processes revolving around forests. It has been launched in 1997 by World Resource Institute (Global Forest Watch, 2002).

At the time of writing Global forest watch offers a lot of data sets, statistics and information related to forests, and also a global map of areas covered by forests and changes in forest cover for the period from 2001 to 2018, developed by remote sensing techniques. On their portal it is effectively presented, there are layers of forest change, land cover, land use, climate and biodiversity. It has information on deforestation, tree cover loss, and gain, as well as deforestation and fire alerts. On their open data portal there are global tree cover loss and tree cover layers at resolution of 30m as well as other useful data available for download. The analysis is done on Landsat 7 data from 600,000 images using Google Earth Engine and supervised learning methods. Tools for analysis and building maps and web gis platforms using their data can also be found there. All can be used for immediate analysis and visualization on whole of Earth. It can be accessed on https://www.globalforestwatch.org/

3.3. FISE

Forest Information System for Europe (FISE) is a hub for the data of classified forests on territory of Europe, for many different species at 1km resolution. It was called for by European Commission and published in 2013. The produced raster maps and diagrams were generated by modeling harmonized presence/absence data sets from many different institutions and sources: European National Forestry Inventories database, Forest Focus/Monitoring data set, BioSoil data set, European Information System on Forest Genetic Resources (EUFGIS), Geo-referenced Database of Genetic Di-versity (GD), as well as harmonized forest cover data sets: Pan European Forest Type Map 2006 (FTM), CORINE Land Cover map 2006 (CLC), ESA GlobCover 2009 (EGC). Density of data from different data sets is visible on Figure 1.

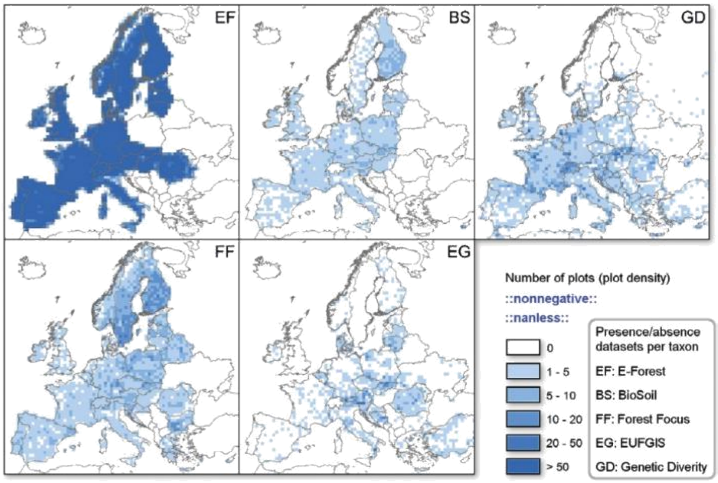

Figure 1. Plot density, computed with a spatial grid of 50 km^2, (LAEA) for the data sets used: European Forest Inventories (EF), BioSoil (BS), Forest Focus (FF), EUFGIS (EG), Genetic Diversity (GD). Among other constraints, the array-based semantics of each harmonised data set expects the corresponding geospatial records to have nonnegative values (::nonnegative::), after the removal of outliers, highly uncertain or missing data (::nanless::), all considered as not available information (San-Miguel-Ayanz, 2016).

European National Forestry Inventories provided approximately 375,000 sample presence/absence data points with a spatial resolution of 1 km, covering 21 European countries. Forest Focus project was started with a goal of creating a harmonized, broad-based, comprehensive and long-term monitoring of air pollution effects in European forests. It has been carried out by participating countries by creating a systematic networks of observation points for periodic and continuous monitoring. Forest Focus/Monitoring data set covers 30 European Countries with more than 8,600 sample points.

BioSoil was started in response to Forest Focus. It’s aim was to provide harmonised soil and forest biodiversity data. In FISE project Biodiversity data was used, which contains 3,300 sample points in 19 European countries.

European Information System on Forest Genetic Resources (EUFGIS) aims to "maintain, conserve, restore and enhance the biological diversity of forests, including their genetic resources, through sustainable forest management". It has an online portal with 2,500 samples of 98 forest species in 31 European countries.

Geo-referenced Database of Genetic Diversity (GD) consists of sample points of trees that are analyzed in genetic surveys. It is part of EVOLTREE project. It was launched in April 2006 and is financially supported by European Union. It contains geographic information of a limited number of tree species, mostly Pine, Oak, Beech and Ash.

Pan European Forest Type Map 2006 (FTM) is derived from satellite imagery by using Neural Network Clustering algorithm. It has classified broad-leaved and coniferous forests in spatial resolution of 25m and covers 38 European countries.

CORINE Land Cover map 2006 (CLC) was described in a previous chapter.

ESA GlobCover 2009 (EGC) is a set of data describing the global land cover in 22 classes defined by United Nations Land Cover Classification System in 10 arc seconds spatial resolution (about 300m at equator). The Constrained Spatial Multi-Scale Analysis (C-SMFA) was used to process the data. The forest classes are described through percentage range of tree cover.

The model used for generating relative probability of presence (RPP) maps for each taxon/species is Constrained Spatial Multi-Scale Analysis (C- SMFA). Each map is modeled with a spatial frequency analysis of the available field observations. In particular, information on the spatial probability of finding a broad-leaved (or coniferous) tree species has been used as C-SMFA statistic constraint to improve the accuracy of the estimation. On Figure 2 forest densities can be seen. To deal with the huge amounts of data - the big data problem, the geospatial semantic array programming paradigm (GeoSemAP) has been exploited.

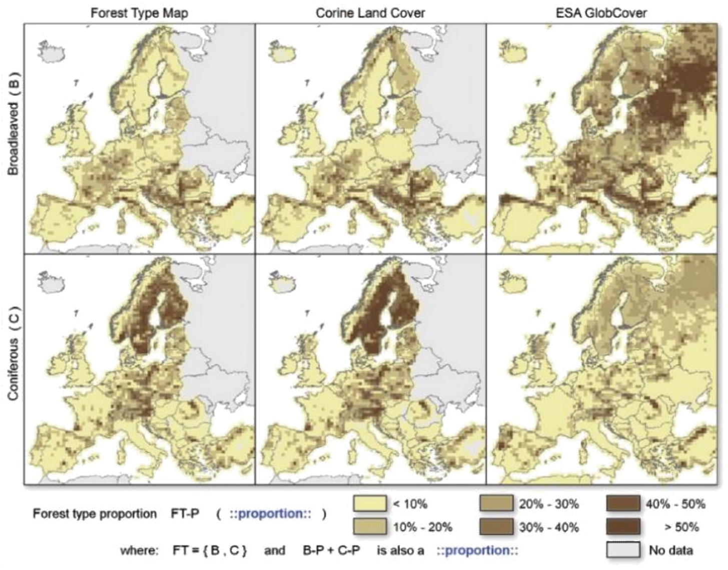

Figure 2. Broadleaved and coniferous forest density, computed with a spatial grid of 50 km^2, (LAEA) for the data sets used: Pan European Forest Type Map 2006 (FTM), CORINE Land Cover map 2006 (CLC) and ESA GlobCover 2009 (EGC). Among other constraints, the array-based semantics of each harmonised forest density expects the corre-sponding geospatial raster layers to provide the proportion of forest cover (::proportion::), i.e. values in [0 1] (San-Miguel-Ayanz, 2016).

Maximum Habitat Suitability (MHS) maps were generated using Relative Distance Similarity model (RDS-MHS). High values represent highly suitable areas for the selected species to survive. In these areas bioclimatic conditions are very similar to fields in which the species has been confirmed. Big Data problem was also present in this analysis, and mitigated by GeoSemAP approach. The model highlights highly suitable and unsuitable areas in Europe. This is not as easily obtained with classical approaches based instead on the average habitat suitability (San-Miguel-Ayanz, 2016).

The data sets can be accessed on https://data.jrc.ec.europa.eu/collection/fise.

4. Discussion

It is argued that all the distributional databases are spatially biased, because of uneven sampling, between regions and countries. For example the study has been done on common eurasian butterfly species *Aglais urticae* which warns of automated applications of species distribution models, because the decline of quality constituted by spatial sampling bias went unnoticed while expert evaluation warned of this (Beck et al. 2014).

It is common practice that primary data are checked for quality before use. This might be a lot of time-consuming work to do, if done for each study individually, so integration of biodiversity platforms and standardization in the field would be the most beneficial from the aspect of biogeographic modeling. This is also a slow process, followed by setting standards along the way and adapting the existing data and integrating with different platforms one by one. Since scientists do not have a direct influence on integration or on the speed of integration of global biodiversity platforms, development of data integration and gathering pipelines for independent studies would make it easier to generate periodically forest species raster maps or studies.

Merging different data sets can be very complicated, time demanding process. In some cases methodology needs to be developed to deal with particular circumstances. The data can be missing, can have missing or out of range values, be wrongly positioned or it can even be overlapping, the same specimen can be found in multiple data sets. Samples are often (especially in modern, phone users generated data points) distributed in easily reachable areas, while more inaccessible areas have few or no samples at all, which has very significant impact on the quality of the final result. The resolution at which the data is to be processed has to be evaluated depending on the type of analysis and amount and density of the available data (Stockwell and Peterson, 2003). On top of all the practical implications, the processing or even searching of large amounts of different data formats (points, polygons, rasters, etc.) can be very demanding on the resources. To deal with this the software solutions to be implemented and hardware resources used in processing should be carefully picked for efficient processing, which can have an impact of several orders of magnitude in processing time (Zhang, 2012).

Biogeographic modeling has three main approaches: descriptive - which try to get an insight into parameters of the environment or maybe biotic interactions of the species, explanatory – try to validate the hypothesis by using a priori knowledge of a system and predictive models – try to obtain the best fit, trying to suggest the optimal environment for the species or for example other locations where it might be found (Guisan and Zimmermann, 2000).

Table 1. The comparison of free open source biodiversity data source platforms.

Primary data sources			
Platform	**No. of records**	**No. od data sets**	**No. of species**
GBIF	1,349,222,286	47,005	4,587,583
iNaturalist	28,138,091	/	239,954
BIEN	206,241,288	/	485,902
RainBio	610,117	/	25,356
DryFlor	208,324	1,600	4,660
Atlas of Living Australia	86,680,052	11,714	654,187
Encyclopedia of Life	/	/	1,300,000+
Secondary data sources			
Platform	**Coverage**	**Products**	**Period**
Copernicus	European and Global	14+	daily to yearly
Global Forest Watch	Global	4+	yearly
FISE	European	39	one time

Modeling can save time and it helps significantly to see the patterns in nature without the need to sample each specimen. Although for good models data must be of proper quality following the data sampling and processing methodology according to the purpose of the models. Today inferred data is becoming more common, and is being used in studies. Also the inferred data must be carefully used, because if something was missed in the modeling, the derived data might be misrepresenting the real state.

It can be seen from the previous chapters that most of the described primary sources of data are using Darwin Core standards, which allows for easy merging of the data sets. Still even though this is very much possible attention must be payed to data used and to the approach of merging.

Currently, as it can be seen in the Table 1, the biggest primary biodiversity data source platform is GBIF. It also integrated some of the other biodiversity platforms data, and offers tools for searching, publishing, referencing and discussing the data sets. It is an excellent source of primary biodiversity data of all kinds for ecological modeling.

From land cover use maps, concerning the forest species, FISE seems as the most complete forest species data set up to date, using the complete information from which data sets have been generated, combines it, harmonizes and adds more value. It is done on data in 2006, and at resolution of 1km, which presents some limitations. On the other hand, all of the data being listed and methodology already developed, it should be easy to perform such modeling on new data sets at new time points.

Land surface under forests with distinction between broad-leaved and coniferous forests seems to be best represented in Copernicus Forest Type data set at very high resolution of 20m. It is periodically generated and has current state of forests described.

5. Conclusion

The forest biodiversity data sources on global level have been searched, listed and briefly described. Most of the sources are improved through time checkpoints, or continuously expanded. The tendency is to unify all the sources into one large comprehensive source where all the important data can be easily reviewed and gathered.

The data merging has been and it will stay one of the most challenging steps in each modeling endeavor. The sheer diversity of the data available and methods of collection complicate things further. Data often on the way to be published loses many useful additional information, and in the steps of integration between platforms it can often lose more. Even though this is true for the past, extensive work on creation, education and spreading of standards and in automation of integration process in biodiversity platforms promises improvements in description and information offered on data currently being published and in future. Hopefully, more automated attempts of integration between the data sets, not just platforms will commence soon (Guralnick, 2006). This would take the work of data merging out of hands of modelers almost completely. This is already having an effect on modelers. It is becoming easier to get the data, and majority of intellectual effort is being shifted to the models and predictions applied.

This paper points out the importance of the standards in forest monitoring and data collection and sharing of once collected data. It gives a good starting point for online data collection and sharing and preparation for biological and forest modeling. It points to specific global platforms for biological data sharing and to useful secondary data sources. It briefly explains the problems of data selection and preprocessing. Using the platforms presented in this work should help finding data for the decision making and biogeographic modeling, and this should not present an obstacle anymore.

References

1. Akay, A.E., Oğuz, H., Karas, I.R. and Aruga, K. (2009): Using LiDAR technology in forestry activities. Environmental Monitoring and Assessment 151: 117-125.

2. Avram, H.D. (2003): Machine-readable cataloging (MARC) program. Encyclopedia of Library and Information Science 3, p. 1712.
3. Jan, B., Boller, M., Erhardt, A., Schwanghart, W., (2014): Spatial bias in the gbif database and its effect on modeling species' geographic distributions. Ecological Informatics 19: 10-15.
4. Lee, B., Williams, K.J. (2016): Towards a national bio-environmental data facility: experiences from the atlas of living australia. International Journal of Geographical Information Science 30: 108-125.
5. Trevor, B.H., Williams, K.J., Belbin, L. (2012): Developing biodiverse plantings suitable for changing climatic conditions 2: Using the atlas of living australia. Ecological Management & Restoration 13: 274-281.
6. Bossard, M., Feranec, J., Otahel, J. (2000): Corine land cover technical guide - Addendum 2000. EEA, Copenhagen.
7. Canhos, V.P., de Souza, S., De Giovanni, R., Canhos, D.A.L. (2004): Global Biodiversity Informatics: setting the scene for a "new world" of ecological forecasting. Biodiversity Informatics 1: 1-13.
8. Catapano, T., Hobern, D., Lapp, H., Morris, R.A., Morrision, N., Noy, N., Schildhauer, M., Thau, D. (2001): Recommendations for the Use of knowledge organisation systems by GBIF. Copenhagen: Global Biodiversity Information Facility, 49 p.
9. Dauby, G., Zaiss, R., Blach-Overgaard, A., Catarino, L., Damen, T., Deblauwe, V., Dessein, S., Dransfield, J., Droissart, V., Duarte, M.C., Engledow, H., Fadeur, G., Figueira, R., Gereau, R.E., Hardy, O.J., Harris, D.J., de Heij, J., Janssens, S., Klomberg, Y., Ley, A.C., Mackinder, B.A., Meerts, P., van de Poel, J.L., Sonké, B., Sosef, M.S.M., Stévart, T., Stoffelen, P., Svenning, J-C., Sepulchre, P., van der Burgt, X., Wieringa, J.J., Couvreur, T.L.P. (2016): RAINBIO: a mega-database of tropical African vascular plants distributions. PhytoKeys 74: 1-18.
10. Brian, E.J., Condit, R., Peet, R.K., Schildhauer, M., Thiers, B.M. (2016): Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. PeerJ Preprints 4: e2615v2.
11. Feinsinger, P. (2001): Designing field studies for biodiversity conservation. Island Press.
12. GBIF.org (2019). Gbif home page. available from: https://www.gbif.org [Accessed: 20 September 2019].
13. Global Forest Watch (2012). Global forest watch. World Resources Institute, Washington, DC Available from http://www.globalforestwatch.org [Accessed: 29 March 2019].
14. Guisan, A., Zimmermann, N.E. (2000): Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.
15. Guisan, A., Thuiller, W., Zimmermann, N.E. (2017): Habitat suitability and distribution models: with applications in R. Cambridge University Press
16. Guralnick, R.P., Wieczorek, J., Beaman, R., Hijmans, R.J. BioGeomancer Working Group (2006): BioGeomancer: automated georeferencing to map the world's biodiversity data. PLoS biology 4: p.e381.
17. Hjerppe, R., Olander, B. (1989): Cataloging and expert systems: AACR2 as a knowledge base. Journal of the American Society for Information Science 40: 27-44.
18. Jørgensen, S.E., Chon, T.S., Recknagel, F. (2009): Handbook of ecological modelling and informatics. Wit Press.
19. Jovanović, D., Govedarica, M., Sabo, F., Bugarinović, Ž., Novović, O., Beker, T., Lauter, M. (2015): Land cover change detection by using remotesensing: A case study of Zlatibor (Serbia). Geographica Pannonica 19: 162-173.

20. Banda-R, K., Delgado-Salinas, A., Dexter, K.G., Linares-Palomino, R., Oliveira-Filho, A., Prado, D., Pullan, M., Quintana, C., Riina, R., et al. (2016): Plant diversity patterns and their conservation implications in neotropical dry forests. Science 353: 1383-1387.
21. Parr, C., Wilson, N., Leary, P., Schulz, K., Lans, K., Walley, L., Hammock, J., Goddard, A., Rice, J., Studer, M., Holmes, J., Corrigan, Jr. R. (2014): The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. Biodiversity Data Journal 2: e1079.
22. Gernot, R., Buzzo, G., Berndt, F., Langanke, T., Herrman, D. (2017): Copernicus land monitoring service high resolution layer forest: Product specifications document European Environmental Agancy.
23. San-Miguel-Ayanz, J., de Rigo, D., Caudullo G., Houston Durrant, T., Mauri, A., Tinner, W., Ballian, D., Beck, P., Birks, H. J. B., Eaton, E., Enescu, C. M., Pasta, S., Popescu, I., Ravazzi, C., Welk, E., Abad Viñas, R., Azevedo, J. C., Barbati, A., Barredo, J. I., Benham, S. E., Boca, R., Bosco, C., Caldeira, M. C., Cerasoli, S., Chirici, G., Cierjacks, A., Conedera, M., Da Ronch, F., Di Leo, M., García-Viñas, J. I., Gastón González, A., Giannetti, F., Guerrero Hue, N., Guerrero Maldonado, N., López, M. J., Jonsson, R., Krebs, P., Magni, D., Mubareka, S., Mulhern, G., Nieto Quintano, P., Oliveira, S., Pereira, J. S., Pividori, M., Räty, M., Rinaldi, F., Saura, S., Sikkema, R., Sitzia, T., Strona, G., Vidal, C., Vilar, L., Zecchin, B. (2016): European Atlas of Forest Tree Species. European Commission. ISBN: 978-92-79-36740-3.
24. Shao, K.T., Peng, C.I., Yen, E., Lai, K.C., Wang, M.C., Lin, J., Lee, H., Alan, Y., Chen, S.Y. (2007): Integration of biodiversity databases in Taiwan and linkage to global databases. Data Science Journal 6: S2-S10.
25. Stockwell, D., Peterson, A.T. (2003): Comparison of resolution of methods used in mapping biodiversity patterns from point-occurrence data. Ecological Indicators 3(3): 213-221.
26. Tang, L., Shao, G. (2015): Drone remote sensing for forestry research and practices. Journal of Forestry Research 26: 791-797.
27. Tomppo, E., Andersson, K. (2008): Technical review of FAO's approach and methods for National Forest Monitoring and Assessment (NFMA). Forestry Department, FAO.
28. Weibel, S. (1997): The Dublin Core: a simple content description model for electronic resources. Bulletin of the American Society for Information Science and Technology 24: 9-11.
29. Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D. (2012): Darwin Core: an evolving community-developed biodiversity data standard. PloS one 7: e29715.
30. Edward, W.O. (2003): The encyclopedia of life. Trends in Ecology & Evolution, 18: 77-80.
31. Zhang, J. (2012): A high-performance web-based information system for publishing large-scale species range maps in support of biodiversity studies. Ecological Informatics 8: 68-77.