



European Vegetation Archive

Data Request Form



To obtain data from the European Vegetation Archive (EVA), including the ReSurveyEurope Database, please first enquire the EVA database administrator Ilona Knollová (ikuzel@sci.muni.cz) whether the data that meet your needs are available. If they are, please fill in the form below and submit it to Ilona or another member of the EVA Coordinating Board (or ReSurveyEurope Board if you ask for data from the ReSurveyEurope Database).

- Applicant's name:

Burcu KARAGOL AYAN

- Applicant's institute:

Google

- Applicant's institutional address:

1600 Amphitheatre Pkwy, Mountain View, CA 94043, USA

- Applicant's e-mail:

burcuka@google.com

- Project title:

ML approaches for developing a global terrestrial species distribution model

- Are you asking for core EVA data (non-repeated vegetation surveys) or for ReSurveyEurope data (repeated vegetation surveys)?

Core EVA data

- Brief description of the aims and methods of the study:

Our study goal is to develop new best-in-class, ML-powered approaches for species distribution modeling. To train our models, we will use observations from the Global Biodiversity Information Facility, and aim to use the EVA dataset as a high quality evaluation dataset.

To advance the understanding, conservation and restoration of biodiversity, we aim to open source the results of our research in collaboration with our partners at UNEP-WCMC, and to publish the results of our research, for which we may want to include small excerpts of the data as an example, but would give attribution to your dataset.

We will not use the data directly in any commercial product or service.



European Vegetation Archive Data Request Form



- Will someone else be involved in data editing or analysis in addition to the applicant?

Researchers and engineers of the sustainability and biodiversity programs at Google (e.g., Drew Purves, Christine Kaiser-Chen, and so on)

- Estimated time of delivery of results (e.g., manuscript submission):

We plan to write papers about our work; our first submission deadline is May 15, 2025

- Geographic area needed (e.g., countries or range of geographic coordinates):

Europe and adjacent areas. Specifically the following countries:

Andorra
Austria
Belgium
Bosnia and Herzegovina
Bulgaria
Croatia
Czech Republic
Denmark
France
Germany
Greece
Hungary
Ireland
Italy
Latvia
Luxembourg
Monaco
Montenegro
Netherlands
Norway
Poland
Portugal
Romania
Serbia
Slovakia
Slovenia
Spain
Switzerland
The former Yugoslav Republic of Macedonia

- Do you need plots to be georeferenced? If so, what is the minimum accuracy of plot location (in metres or kilometres) needed for your project?

Plots need to be georeferenced (they should have degrees of latitude and longitude) but



European Vegetation Archive Data Request Form



there is not a minimum accuracy of plot location needed.

- Vegetation types needed (syntaxa):

All of European flora

- Other data selection criteria:

To assess our species distribution modeling approach, we plan to use GeoPlant train/test data split (<https://arxiv.org/pdf/2408.13928> and <https://www.kaggle.com/datasets/picekl/geoplant/data>). We are interested in the region, taxa, georeferenced plots and uncertainty in their datasets.

We would like to request the data necessary to deanonymize the species IDs used in the GeoPlant dataset. This will allow us to evaluate the impact of adding species traits in species distribution models. We would also like to request any plant trait data that is available for this purpose.

[Lukas Picek](#) of the University of West Bohemia, lead author of the GeoPlant study, is aware and supportive of our research plans for this data.

- Envisaged publications:

Conference proceeding for the Thirty-Ninth Annual Conference on Neural Information Processing Systems 2025 (NeurIPS)

- Data deposition.** Some journals require data used for the analysis to be stored in a public repository to ensure the repeatability of the analysis. According to EVA Rules, you are not allowed to store the original vegetation-plot data obtained from EVA. However, if you plan to publish in such a journal, you may deposit a reduced EVA-derived dataset that (1) would make it possible to repeat the analysis published in the paper and (2) does not contain any information not used in the analysis. For example, such a dataset can contain only a subset of species (e.g., only angiosperms or only neophytes), or replace species names with codes, or replace species cover values with presences/absences, or remove all the header data, or replace the exact plot coordinates by coarse grid-cell coordinates etc. If you plan to deposit reduced information from vegetation plots, please describe here what might be deposited. If the project developed so that you needed to deposit more information than specified here, you would need to ask specific permission from the Custodians of the EVA databases used in your analysis before the dataset is deposited.

We do not plan to deposit the source data. We may publish representative example results from our evals.

- Plant trait data from the TRY consortium.** If you plan to combine your analysis of vegetation-plot data with plant trait data, you can also request a dataset of 18 gap-filled traits for a large number of plant taxa prepared by the TRY consortium. These traits include Leaf area, Specific leaf area, Leaf fresh mass, Leaf dry matter content, Leaf C, Leaf N, Leaf P, Leaf N per area, Leaf N:P ratio, Leaf delta15N, Seed mass, Seed length, Seed number per reproductive unit, Dispersal unit length, Plant height, Stem specific density,



European Vegetation Archive Data Request Form



Stem conduit density, and Conduit element length. This dataset can be provided to you by the EVA manager together with the vegetation-plot data. If you use this dataset, you must inform about your project the TRY data contributors who might be potentially interested and invite them as potential co-authors, assuming they will make an intellectual contribution to your paper. The list of the TRY data contributors will be sent to you together with the gap-filled trait dataset.

We are interested in learning more about the TRY dataset because of potential applicability for model training; is there any additional information you can share about the number species covered?

- Specification of the co-authorship arrangements in publications based on the requested data. Note that the EVA Rules recommend that co-authorship is offered to a representative of each database providing data that are particularly important for the project (e.g., a relatively large proportion of the final dataset used in the analyses or data from unique vegetation types or under-represented geographic areas). This database representative should be an expert in the topic of the project (not necessarily the custodian or deputy custodian), and this person should contribute to the project more than just by providing the existing data, e.g. by intellectual contribution to the concept of the paper, preparation of new data, or helping with data analysis, interpretation of the results or writing parts of the paper (see the IAVS Code of Professional Ethics: https://www.iavs.org/page/governance_code-of-professional-ethics). The project leader should enable active participation by regularly informing potential co-authors about the progress of the project from its early stage. The project leader should also make final co-authorship arrangements based on the real input of the individual contributors.

The data will be a part of an existing set for evaluation of large machine learning models for species distribution modeling. We do not plan to use the data in our models and we do not expect any intellectual contributions from the data providers, therefore we plan to “just” acknowledge the database custodians in the expected publications.

- Eligibility of the applicant to receive EVA or ReSurveyEurope data. Specify to which EVA or ReSurveyEurope database the applicant has contributed; if the applicant is not the custodian or deputy custodian of an EVA or ReSurveyEurope database, give a name of a custodian or deputy custodian who supports this data request.

Not contributed. Milan Chytrý, the Custodian of the Czech National Phytosociological Database, recommends this project.

- I agree with the terms of EVA Data Property and Governance Rules as approved on 26 May 2012 (<http://euroveg.org/download/eva-rules.pdf>).
- If I ask for ReSurveyEurope data, I agree with the terms of ReSurveyEurope Data Property and Governance Rules as approved on 6 April 2022 (<http://euroveg.org/download/resurveyeurope-rules.pdf>).
- In any result obtained based on EVA core data (non-repeated vegetation surveys), I will cite the EVA report article (Chytrý et al. 2016; <https://doi.org/10.1111/avsc.12191>). In any result obtained based on the ReSurveyEurope data (repeated vegetation surveys), I will cite the ReSurveyEurope report article as soon as it is published. In addition, I will cite individual source databases used in my project (if possible, in the list of References; if not possible, at least as a list of databases in the electronic supplementary material).



European Vegetation Archive Data Request Form



- If I ask for the plant trait data from TRY, I agree to invite to my project the TRY data contributors following the list received from the EVA database manager.

California, 6 April 2025

Burcu Karagol Ayan