

Unlocking Slovakia's Biodiversity Through Data Publishing

Programme: CESP

Project ID: CESP2023-005

Project page: <https://www.gbif.org/project/CESP2023-005/unlocking-slovakias-biodiversity-through-data-publishing>

Project lead organization: Plant Science and Biodiversity Centre, Slovak Academy of Sciences

GBIF Participant Represented: GBIF Slovakia

Project implementation period: 1/10/2023 - 30/9/2024

Report approved: 1/5/2024

Declaration: The information provided in this report has been provided by and is the responsibility of the project lead.

Narrative Midterm report

Executive Summary

The project aims to mobilize biodiversity data that has not yet been published in the GBIF. The database of distribution data records PLADIAS-SK that will serve as a source of data for publishing is fully functional, containing as a test case more than 5000 distribution records of vascular plants from the area of Slovakia. The pipeline for import of data from the phytocoenological database TURBO(VEG) Slovakia into PLADIAS-SK was established. It includes "cleaning of the data", which means the use of names accepted in the taxonomic backbone of PLADIAS-SK, checking the correctness of geographical coordinates, proper administrative units of the country, etc. The database is already connected to the GBIF portal. The test case of 5000 distribution data records from the PLADIAS-SK database is visible via GBIF portal. The migration of new data from TURBO(VEG) Slovakia into PLADIAS-SK is on the way. The project will also provide training to individuals from different organizations in data carpentry skills required for biodiversity data publication. Participants will learn how to organize and format data, create metadata, and use the Integrated Publishing Toolkit (IPT) to publish data in GBIF. The preparation of the workshop is on the way. The date was already fixed for 8.-10. May 2024. We already booked accommodation and seminary room at Stara Lesna in the Tatry Mts. We are just receiving applications from individual participants.

Report on Activities

Activity progress summary

Activity "Workshop": the preparation of the workshop is on the way. The date was already fixed for 8.-10. May 2024. We already booked accommodation and seminary room at Stara Lesna in the Tatry Mts. We are just receiving applications from individual participants.

Activity "PLADIAS-SK": this activity is almost completed. The database of distribution data records PLADIAS-SK is fully functional, containing as a test case more than 5000 distribution records of vascular plants from the area of Slovakia. The pipeline of import of data from the phytocoenological database TURBOVEG Slovakia into PLADIAS-SK was established. It includes "cleaning of the data", which means the use of names accepted in the taxonomic backbone of PLADIAS-SK, checking the correctness of geographical coordinates, proper administrative units of the country, etc. The database is, however, not yet accessible and visible of external users.

Activity "TURBO(VEG) data into the PLADIAS-SK": this activity is at its beginning and the migration of TURBO(VEG) data into CMS PLADIAS-SK required fully functional database. As stated above, test data were already migrated and we are working on the preparation of the main portion of data for migration.

Activity "PLADIAS-SK connected to GBIF": this activity is almost completed. We developed a Python application deployed in Docker, aimed at extending the capabilities of the PLADIAS-SK database, a species observation platform originating from Czechia and now expanding into Slovakia. This application interfaces with the PLADIAS-SK's Postgres database to run a specialized query, extracting data into a Darwin Core compliant CSV file. Upon generation, this CSV is uploaded to Zenodo, effectively creating a new dataset version. The IPT then utilizes this CSV, appending metadata and

facilitating its publication on GBIF. It's important to stress the ongoing continuous migration of the occurrence data from the phytocoenological database TURBOVEG Slovakia, expected within the current project, making the currently published dataset a preliminary preview of the pipeline's potential. As more data is migrated and integrated into PLADIAS-SK, the pipeline is designed to automatically process and publish these updates, ensuring a dynamic and evolving dataset reflective of the most current biodiversity records.

It's important to note that migration of previous Institute of Botany (Now Plant Science and Biodiversity Centre, Institute of Botany) Slovak Academy of Sciences legacy dataset (registered under the name DATAFLOS) into PLADIAS-SK and via this platform to the GBIF is expected after finishing the import of the data from TURBOVEG phytocoenological database.

Report on Deliverables

Deliverables progress summary

Data Mobilisation Workshop: the preparation of the workshop is on the way. The date was already fixed for 8.-10. May 2024. We already booked accommodation and seminary room at Stara Lesna in the Tatry Mts. We are just receiving applications from individual participants.

PLADIAS-SK: the work on this deliverable is almost completed. The database of distribution data records PLADIAS-SK is fully functional, containing as a test case more than 5000 distribution records of vascular plants from the area of Slovakia. The pipeline of import of data from the phytocoenological database TURBOVEG Slovakia into PLADIAS-SK was established. It includes "cleaning of the data", which means the use of names accepted in the taxonomic backbone of PLADIAS-SK, checking the correctness of geographical coordinates, proper administrative units of the country, etc. The database is, however, not yet accessible and visible of external users.

TURBO(VEG) data into the PLADIAS-SK: the work on this deliverable already started. As stated above the first testing dataset of 5000 distribution data records of vascular plants from Slovakia was already migrated into PLADIAS-SK database.

PLADIAS-SK connected to GBIF: also the work on this deliverable is almost completed. Few technical issues remained to be solved.

Communications and visibility

During the first half of the duration of the project we concentrated on functionalities of the CMS - PLADIAS-SK database, on the pipeline for the migration of data from the TURBO(VEG) database of phytocoenological records into PLADIAS-SK and on the implementation of the connection of PLADIAS-SK database. As yet we did not have anything to present publicly, except of the reporting on the progress at the meetings of the Slovak Botanical Society or Slovak Academy of Sciences meetings. As soon as a reasonable portion of data will be migrated into the PLADIAS-SK database and presented also via GBIF portal, we will make several public presentations at meetings and web pages (as expected in the project proposal).

Monitoring and evaluation

Monitoring and evaluation findings

The progress on the project is well on the way. The work is going on in all activities. Deliverables/activities scheduled to be finished in June and July 2024 are almost completed and the last deliverable aimed at completing in September 2024 is in progress. This last deliverable deals with the migration of distribution data on vascular plants of Slovakia from the phytocoenological database TURBO(VEG) into the database PLADIAS-SK and subsequently as one dataset to be published at GBIF portal. Test dataset for this purpose is already available. The data from the database TURBO(VEG) require certain amount of "cleaning", namely update of taxonomy to the PLADIAS-SK taxonomic backbone, checking the correctness of geographical coordinates, linking to the administrative regions, etc. Anyway we believe that we will be able to complete migration of data in time.

The Capacity Support Enhancement Programme (CESP) is funded by GBIF

