Open Forest Data: Digitalizing and building an online repository

Beata Bramorska ‡

‡ Mammal Research Institute of Polish Academy of Sciences, Białowieża, Poland

Corresponding author: Beata Bramorska (bbramorska@ibs.bialowieza.pl)

Abstract

Poland is characterised by a relatively high variety of living organisms attributed to terrestrial and water environments. Currently, close to 57.000 species of living organisms are described that occur in Poland (Symonides 2008), including lowland and mountain species, those attributed to oceanic and continental areas, as well as species from forested and open habitats. Poland comprehensively represents biodiversity of living organisms on a continental scale and thus, is considered to have an important role for biodiversity maintenance.

The Mammal Research Institute of Polish Academy of Sciences (MRI PAS), located in Białowieża Forest, a UNESCO Heritage Site, has been collecting biodiversity data for 90 years. However, a great amount of data gathered over the years, especially old data, is gradually being forgotten and hard to access. Old catalogues and databases have never been digitalized or publicly shared, and not many Polish scientists are aware of the existence of such resources, not to mention the rest of the scientific world.

Recognizing the need for an online, interoperable platform, following FAIR data principles (findable, accessible, interoperable, reusable), where biodiversity and scientific data can be shared, MRI PAS took a lead in creation of an Open Forest Data (OFD) repository. OpenForestData.pl (Fig. 1) is a newly created (2020) digital repository, designed to provide access to natural sciences data and provide scientists with an infrastructure for storing, sharing and archiving their research outcomes. Creating such a platform is a part of an ongoing development of life sciences in Poland, aiming for an open, modern science, where data are published as free-access. OFD also allows for the consolidation of natural science data, enabling the use and processing of shared data, including API (Application Programming Interface) tools. OFD is indexed by the Directory of Open Repositories (OpenDOAR) and Registry of Research Data Repositories (re3data).

The OFD platform is based entirely on reliable, globally recognized open source software: DATAVERSE, an interactive database app which supports sharing, storing, exploration, citation and analysis of scientific data; GEONODE, a content management geospatial system used for storing, publicly sharing and visualising vector and raster layers,

©

.

GRAFANA, a system meant for storing and analysis of metrics and large scale measurement data, as well as visualisation of historical graphs at any time range and analysis for trends; and external tools for database storage (Orthanc) and data visualisation (Orthanc plugin Osimis Web Viewer and Online 3D Viewer (https://3dviewer.net/), which were integrated with the system mechanism of Dataverse. Furthermore, according to the need for specimen description, Darwin Core (Wieczorek et al. 2012) metadata schema was decided to be the most suitable for specimen and collections description and mapped into a Dataverse additional metadata block. The use of Darwin Core is based on the same file format, the Darwin Core Archive (DwC-A) which allows for sharing data using common terminology and provides the possibility for easy evaluation and comparison of biodiversity datasets. It allows the contributors to OFD to optionally choose Darwin Core for object descriptions making it possible to share biodiversity datasets in a standardized way for users to download, analyse and compare.

Currently, OFD stores more than 10.000 datasets and objects from the collections of Mammal Research Institute of Polish Academy of Sciences and Forest Science Institute of Białystok University of Technology. The objects from natural collections were digitalized, described, catalogued and made public in free-access. OFD manages seven types of collection materials:

1. 3D and 2D scans of specimen in Herbarium, Fungarium, Insect and Mammal Collections,
2. images from microscopes (including stereoscopic and scanning electron microscopes),
3. morphometric measurements,
4. computed tomography and microtomography scans in Mammal Collection,
5. mammal telemetry data,
6. satellite imagery, geospatial climatic and environmental data,
7. georeferenced historical maps.

In the OFD repository, researchers have the possibility to share data in standardized way, which nowadays is often a requirement during the publishing process of a scientific article. Beside scientists, OFD is designed to be open and free for students and specialists in nature protection, but also for officials, foresters and nature enthusiasts. Creation of the OFD repository supports the development of citizen science in Poland, increases visibility and access to published data, improves scientific collaboration, exchange and reuse of data within and across borders.

Keywords

digital repository, biodiversity, Darwin Core, Darwin Core Archive, open software

Presenting author

Beata Bramorska

Presented at

TDWG 2021

Funding program

This project received partial funding from Operational Programme Digital Poland.

Grant title

e-Puszcza. Data repository for natural scientific data of Podlasie Voivodeship, Poland *(Operational Programme Digital Poland)*

Hosting institution

Mammal Research Institute of Polish Academy of Sciences

Conflicts of interest

References

- Symonides E (2008) Ochrona przyrody. [Nature conservation]. Wydawnictwo Naukowe PWN [In Polish].
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D, et al. (2012) Darwin Core: an evolving community-developed biodiversity data standard. Plos one 7 (1). https://doi.org/10.1371/journal.pone.0029715



Figure 1.

Open Forest Data logo.