EUdaphobase – European Soil-Biology Data Warehouse for Soil Protection

David J. Russell[1*], *Paul Henning Krogh*[2]

[1*]david.russell@senckenberg.de, Senckenberg Museum of Natural History Görlitz, 02826 Görlitz, Germany
[2]Aarhus University, Department of Bioscience, 8600 Silkeborg, Denmark

Abstract summary

Assessment of soil biodiversity, i.e., within the context of national and EU directives and policies, must be based on proper baseline and threshold values as well as reliable tools derived from existing supra-regional data on species' occurrences. Throughout Europe, abundant soil biodiversity information exists, whose common potential has yet to be explored or collated together. The COST Action EUdaphobase is therefore developing an existing soil-biodiversity data infrastructure into a pan-European soil-biodiversity data warehouse, establishing a comprehensive knowledge- and database of high scientific and data-management quality for Europe.

This data warehouse will not only aggregate data on species distributions throughout Europe (like other data repositories), but will combine this data with accompanying environmental metadata and species' functional traits, providing a vastly increased value of data re-use for ecological evaluations. Based on the needs of stakeholders and end users, the Action will evaluate and test modern biodiversity modelling approaches for their implementation in semi-automated decision-support tools for soil-biodiversity assessments. A coordinated approach towards assessment of soil health in terms of soil biodiversity and ecosystem services will ensure more efficient and knowledge-based assessment of soil biodiversity, quality and health.

Keywords: Soil biodiversity, Europe, data harmonisation, data warehouse, FAIR principles, data re-use, assessment tools

Introduction, scope and main objectives

European authorities and stakeholders urgently need reliable tools for monitoring and evaluating the environmental condition of soils, particularly within policy assessment in context of national and EU directives. Many soil functions leading to ecosystem services (ESS) are biotically driven (Turbé *et al.*, 2010), so that soil protection requires coordinated efforts for the evaluation of soil biota throughout Europe (Dunbar *et al.*, 2013). However, without proper baseline data and reliable tools for soil-state assessment, it is currently difficult to efficiently address such goals (Glæsner *et al.*, 2014). Procedures for assessing soil biodiversity as well as establishing baseline values and current states must be based on existing data (Ramirez *et al.*, 2015), preferably accumulated from national or local databanks. International biodiversity databases such as GBIF, PREDICTS or DiSSCo do contain information about species and their distributions, but cannot provide focussed representation of the current status of soil biodiversity. More importantly, they are not operational for ecological assessments or advancing decision-support tools, as they do not include crucial environmental metadata.

The COST Action EUdaphobase is further developing a soil-biodiversity data infrastructure into a pan-European soil-biodiversity data warehouse, establishing a comprehensive database and knowledge portal of high scientific and data-management quality. An aim of the infrastructure is the use of such data in publically available decision-support instruments for effectively addressing EU-level policy goals concerning soil protection. The focus of the Action is on creating the structures, capacities and procedures necessary for developing an open, publically available data warehouse for Europe-wide soil biodiversity data and assessment tools. The goal is to establish such a supra-regional pan-European data and knowledge infrastructure, providing both soil-biological data and instruments to EU and

national institutions, science and private/public stakeholders for understanding, protecting and sustainably managing soils and their biodiversity and functions.

Methodology

The Action's infrastructure is based on the Edaphobase soil-biology data-warehouse platform that includes data from diverse sources (literature, scientific research data, monitoring data, museum collections, etc.) (Burkhardt *et al.* 2014). As opposed to classical data repositories, this data warehouse not only collates, but also harmonizes and integrates heterogeneous data sets, combining biotic occurrence data with georeferenced sites of occurrence and methodological and environmental metadata, and renders all this data reusable for further analyses.

The work plan of the Action follows the basic logical model of information flow from data import by providers, through data curation and harmonization in the data warehouse, to data queries and analyses by end users of data (Fig. 1). To ensure highest possible data congruence and data-provider acceptance, the Action is agreeing upon standardized terminologies, implementing international standards for data harmonisation (i.e., AgroVoc standards; EUNIS and CORINE systems, etc.). A first key activity has been the development of an easy-to-use, flexible tool for data upload, harmonization and import into the data platform, allowing data providers to upload their data and metadata as is, without having to reorganize it to fit standard data structures. Data quality-control procedures developed for national data platforms are being expanded into international protocols and an international review board for taxonomic and technical quality control will be established.

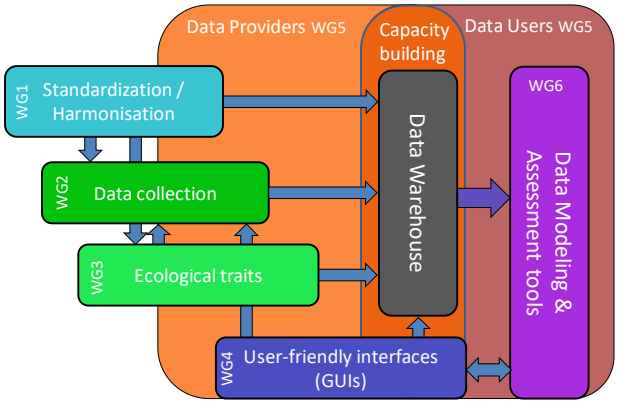

Figure 1: Work plan of the COST Action EUdaphobase.

A focal approach is to include morphological, response and effect trait data for soil-dwelling species collated from Action participants' national databases and integrated into the EUdaphobase data infrastructure. The goal is to combine such trait data with observational data on species distributions as well as with indispensable environmental metadata to gain insight into site-specific functional relationships in soils. Available European, national and remote sensing (i.e. satellite) data will be identified that can fill gaps in environmental metadata.

Innovative procedures to operationalize assessments of the state of soil biodiversity will be developed. Specific Action workgroups will work with regional, national and European stakeholders to identify their data-use and -analysis needs. Based on this, specific analytical tools will be developed for applied uses of policy, management and regulatory agencies. These open-access tools of the data warehouse will recognize and visualize (i.e. on maps) the contribution of soil biodiversity to soil functioning related to soil type, use and management practices as well as determine and delineate soil ESS, baselines and set the basis for forecasting changes thereof.

Results

The EUdaphobase consortium presently includes over 70 participants from ca. 30 pan-European countries. The EUdaphobase infrastructure is based on the Edaphobase platform, which uses developed methodologies for integrating biodiversity and environmental data from diverse sources, such as the Essential Biodiversity Variables (Kissling *et al.*, 2018), the INSPIRE guidelines for soil metadata (https://inspire.ec.europa.eu/), DataCite metadata standards for DOI data citations or the DarwinCore tool for biodiversity-data exchange (Wieczorek *et al.*, 2012). Edaphobase implements the FAIR principles (Findable, Accessible, Interoperable, Reusable; Wilkinson *et al.*, 2016) and offers DOIs for data publication.

Edaphobase can currently accept over 600 different variables relating to data source, taxa, sites, soil properties, climate, methodologies etc. The minimal data set has been defined as taxonomic names ("species"), geo-referenced site information and sampling dates ("what", "where", "when"), whereby recommended are furthermore (meta)data on habitat types, land-use and soil parameters as well as sampling and determination methodologies and persons "under what environmental conditions", ("how" and "by whom"). A data policy and data-sharing agreement regulates open access while ensuring the intellectual property rights (IPRs) of data providers as well as implementing recent European data-protection legislation.

Current data-import software can import text, Excel or Access files into the data-warehouse infrastructure. After registering data-provider information, the software records metadata on the data source, the included sites and environmental parameters. Subsequently, the provider's data is mapped to the Edaphobase data structures, converting nomenclatures and vocabularies where necessary. Finally, the software performs basic (pre-import) quality controls. To ensure high-quality data in the infrastructure, further (peri-import) manual quality-control procedures are being tested, and the data provider can (post-import) control the data in the data warehouse before it is opened to public access.

APIs are currently being planned to network Edaphobase with existing trait databases, e.g. BETSI (Hedde *et al.*, 2012) and EcoTaxonomy (Potapov *et al.*, 2020), and procedures are being developed to link trait data to species' sites of occurrence. Edaphobase currently offers basic descriptive data-analysis tools, such as distribution maps (differentiable according to species or habitat/environmental parameters), species' niche-space analyses as well as expected species composition for specific site conditions (Hausen *et al.*, 2017), implementing procedures for automatically querying data from the database for statistical analyses and models. More detailed prognoses of soil biodiversity (distribution maps as well as site-specific point scales) based on species-distribution models are currently being developed. Based on these software procedures, the Action is conceiving more advanced tools similar to, or in conjunction with, decision-support tools such as the recent Soil Navigator (Debeljac *et al.*, 2019).

Discussion

Throughout Europe abundant soil biodiversity information exists, whose common potential has yet to be explored. However, a lack of consensus on taxonomic classification or standardised vocabularies renders their interoperability difficult. A number of recent European projects on soil-biodiversity evaluation (i.e., ENVASSO, EcoFINDERS, LANDMARK) offer cautious optimism that a common knowledge base can be achieved for key soil-organism groups. To solve these problems, instead of organising another round of harmonised large-scale and costly sampling campaigns to generate *new* data, the EUdaphobase Action is establishing procedures for collecting, curating, quality-checking and harmonizing *existing* data. Thereby, the Action aims towards a pan-European data *warehouse* (not a *repository*) specifically structured to integrate different datasets for common data re-use and *post-hoc* synthesis (cf. Inmon, 2005).

End users request procedures for evaluating questions of soil quality relevant for their specific domains. Tools, e.g. for determining biodiversity baselines and thresholds, must use available data relating taxa to sites of occurrence and to the environmental conditions of those sites. By focussing on including

environmental metadata, the European data and knowledge warehouse will allow the conception of such evaluation tools for applied uses by policy makers, management and regulatory agencies, consultants (SMEs), NGOs, etc.. Furthermore, the inclusion and combination of species functional traits will allow insight to be gained, through these tools, into site-specific functional relationships in soils and to predict the state of biotically driven ecosystem services.

Acknowledgements

EUdaphobase is funded by the COST Association (EU Framework Programme Horizon 2020) under CA 18327 with further support from the The Danish Center for Environment and Energy (DCE), Aarhus University.

References

Burkhardt, U., Russell, D.J., Buryn, R., Decker, P., Döhler, M., Höfer, H., Lesch, S., Rick, S. Römbke, J., Trog, C., Vorwald, J., Wurst, E. & Xylander, W.E.R. 2014. The Edaphobase Project of GBIF-Germany – A new online soil-zoological data warehouse. *Applied Soil Ecology* 83: 3-12.

Debeljak, M., Trajanov, A., Kuzmanovski, V., Schröder, J., Sandén, T. Spiegle, H. Wall, D.P., *et al.* 2019. A Field-Scale Deciosioin Support System for Assessment and Management of Soil Functions. *Frontiers Env Sci* 7:115, doi:10.3389/fenvs.2019.00115.

Dunbar, M.B., Panagos, P. & Montanarella, L. 2013. European Perspective of Ecosystem Services and Related Policies. *Integr. Environ. Assess. Manag.* 9: 231–236.

Glæsner, N., Helming, K. & de Vries, W. 2014). Do current European policies prevent soil threats and support soil functions? *Sustainability* 6: 9538-9563.

Hausen, J., Scholz-Starke, B., Burkhardt, U., Lesch, S., Rick, R., Russell, D., Roß-Nickoll, M. & Ottermanns, R. 2017. Edaphostat: interactive ecological analysis of soil organism occurrences and preferences from the Edaphobase data warehouse. *Database* 2017: 1-6, doi:10.1093/database/bax080.

Hedde, M., Pey, B., Auclerc, A., Capowiez, Y., *et al.* 2012. BETSI, a complete framework for studying soil invertebrate functional traits., 16th International Colloquium on Soil Zoology, Curitiba, Brazil. doi: 10.13140/2.1.1286.6888.

Inmon, W. 2005. *Building the Data Warehouse*, 4th Edn. John Wiley and Sons, 576 pp.

Kissling, W.D., Ahumada, J.A., Bowser, A., Fernandez, M., *et al.* 2018. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biol. Rev. 93: 600–625.

Potapov, A., Sandmann D. & Scheu, S. 2019. Ecotaxonomy virtual research environment – traits and species. Available at http://ecotaxonomy.org. Accessed at 1/31/2020.

Ramirez, K., Döring, M., Eisenhauer, N., Gardi, C., *et al.* 2015. Toward a global platform for linking soil biodiversity data. *Frontiers Ecol. Evol.* 3: 91, doi: 10.3389/fevo.2015.00091

Turbé, A., De Toni, A., Benito, P., Lavelle, P., *et al.* 2010. *Soil biodiversity: functions, threats and tools for policy makers*. Bio Intelligence Service, Report for European Commission, 249 pp.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., *et al.* 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7(1):e29715. doi:10.1371/journal.pone.0029715.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., *et al.* 2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018 doi:10.1038/sdata.2016.18.