



Growing  
**ideas**  
through  
**networks**

# CA-18237 EUdaphobase

European Soil-Biology Data Warehouse for Soil Protection

**edapho**  
**base** Edaphobase Data Policy - Overview



## Data policy

Balancing open access (FAIR) & intellectual property rights (IPRs)

### *Policy regarding data*

- “Regulates”:
  - „data provision“ to the database
  - „data usage“ from the database
- *Determined in/with European consensus/majority*<sup>1</sup>

## Data policy

Balancing open access (FAIR) & intellectual property rights (IPRs)

### Ethics of Data sharing

#### Central concept = „Consent“

e.g., Skovgaard et al. 2019  
Boté & Tèrmins 2019

**Ideal:** Individual, *explicit*, informed consent  
→ difficult to automate; not “open access”

**Edaphobase:** General, *implicit*, implied consent  
→ dynamic (i.e. can be changed)  
→ for restricted data, individual explicit consent possible

# Data policy

from data-provider perspective → intellectual property rights (IPRs)

**Edaphobase differentiates between:**

- “Data owners”
- “Data providers”

## **Data owners:**

- Person (institution) that has collected and compiled (= generated) the original data
- Has ownership (copyright) rights to the data and metadata
- Can be author of a publication, PI of a project, person identifying collected specimens

## **Data provider<sup>1</sup>:**

- Person uploading data to Edaphobase
- The contact person for potential questions from Edaphobase
- Not necessarily, but can be the same as the data owner

<sup>1</sup> In the following, where "data provider" is mentioned, applies to both the data provider and the data owner(s)

# Data policy

from data-provider perspective → intellectual property rights (IPRs)

**Edaphobase:** publicly available data warehouse  
data = *generally* open access

But: **3 Options** of public data access (chosen by data provider)

## (1) Default = Open Access:

data providers: agree to a data-sharing agreement  
→ acknowledge the data policy  
→ agree to public re-use (incl. data download)  
= “*broad, implicit consent*”

≠ “open data” !!

Due to: a) providers retain control (i.e., retract / change data, access options)  
b) further restriction options

# Data policy

from data-provider perspective → intellectual property rights (IPRs)

## (2) Embargo

public data access temporarily restricted (i.e., to allow publication by provider)  
Standard: **3 years**; can be withdrawn or extended

## (3) Restricted access („anonymization“)

public access of partial (“sensitive”) data permanently restricted

- Geo-coordinates = non exact ;
- data allowing identification of sites or sampling design not publicly available:
  - Site, Plot names, abbreviations
  - Habitat description (text fields)
  - Sample numbers, volumes etc.
  - Taxon quantities (counts, densities, etc.)
  - Associated remarks or original quotations

# Data policy

from data-provider perspective → intellectual property rights (IPRs)

## (2) Embargo

public data access *temporarily* restricted (i.e., to allow publication by provider)  
Standard: **3 years**; can be withdrawn or extended

## (3) Restricted access („anonymization“)

public a

- G
  - da
- Such data *cannot* be publicly accessed or downloaded  
→ Shown as: “*exist, but not publicly accessible*”  
→ is used in internal data-analysis tools <sup>1</sup>
  - can be released by the data provider to a specific data user  
(= “*personal, explicit, informed consent*”)

- Associated remarks or original quotations

# Data policy

from data-provider perspective → intellectual property rights (IPRs)

## 1 What does „used in internal tools“ mean?

E.g.:

Mapping sites of occurrence

Data aggregations

in matrices (contingency tables), e.g. Species X Country

EdaphoStat tools

Frequency distributions along, e.g., habitat types,  
Density regressions along gradients (e.g., soil pH)  
etc.

- Associated remarks or original quotations



# Data policy

from data-provider perspective → intellectual property rights (IPRs)

## Further option settings:

- Share data with linked databases (e.g., GBIF)  
→ yes/no answers
- Request a DOI for the data set (*registered with DataCite*)  
→ permanent! Additional metadata required  
(*DOIs only given to “data owners”*)
- Metadata to „understand“ the data  
→ „Scope“ (which?)  
→ Methods, environmental metadata, etc.

# Data policy

from data-user perspective → “Open Access”

data = *generally* open access, downloadable only by registered users  
*exception: embargoed or restricted data (unless individual permission)*

## First step: terms of use:

- Credit the data providers
- If granted access to restricted data sets, not allowed to further disseminate
- Not allowed to use data for commercial use or profit without obtaining written permission from the data providers

**This must be confirmed in Edaphobase before data can be downloaded!!**  
**(Downloads are NOT logged!)**

*Note:*

→ **Nothing** can be enforced!

Provides legal background for regress claims by data providers!

# Data policy

from data-user perspective → “Open Access”

## Citing data providers

### ➤ Co-Authorship

- i.e., for “substantial” data (i.e. ca. >5%) or unique or novel data sets
- Contact information NOT provided publically, must be requested

### ➤ Citation

- Using, e.g., DOI, literature citation (if data derived from publication), citation suggestion for data sets

### ➤ Acknowledgements

- All other cases

# Data policy

from data-user perspective → “Open Access”

## ➤ Data downloads in the form of a „package“

Including:

- Terms of Use (= the relevant sections of the data policy)
- Query Metadata (*for reproduction, methods sections, etc.*)
  - Date of data access
  - Which filters/grouping variables were used
  - How many records, taxa, locations, sources were found
  - List of most frequent sources
- The data itself
  - in flat tables: each data record begins with the source!

# Data policy

from data-user perspective → “Open Access”

➤ **Data downloads in the form of a „package“**

Including:

- Terms of Use (= the relevant sections of the data policy)
- Query Metadata (*for reproduction, methods sections, etc.*)
  - Date of data access
  - Which filters/grouping variables were used
  - How many records, taxa, locations, sources were found
  - List of most frequent sources
- The data itself

**first columns (of a flat download table) will concern the source**  
(e.g., publication, DOI, project, collection, name of author / PI / data provider)

→ this to be used in citing data providers

# Data policy

*Further aspects*



SENCKENBERG  
world of biodiversity

## All aspects of how data are handled and the use of the data warehouse

### **The data policy further explains and regulates:**

- Rights & responsibilities of data providers and of Edaphobase itself
- Privacy statement (how personal data are handled)
- Data sharing with networked databases
- Terms of use of the browser-based Data-Query Portal
- Etc.

→ *These are not shown here; please see the data policy itself*