**Legend to datasets**

1. **data\_table:**

**site** Unique study site ID

**transect** Transect number within site

**interval** sampling interval

**year** Year in which the data were recorded

**group** Taxonomic group

**species** Species level ID

**abundance** Abundance of species (in collection\_method is malaise\_trap, abundances are presence / absence)

**feeding\_guild** feeding guild the species belongs to (used to differentiate between prdators and granivores in carabid beetles and birds)

**comparable\_between\_years** Logic, if TRUE than data are directly comparable between both years of recording (same methodology and sampling effort, same point in the phenology)

**additional\_transects\_in\_large\_sites** Logic, if TRUE than transects were only performed in large sites and are not included in all models investigating species area relationships

**species\_level\_ID** Logic, if FALSE, only morphospecies or genus could be identified

**collection\_method** Method of collecting

**presence\_absence\_data** Logic, if TRUE, data is only presence / absence and not abundance data

1. **environment\_table:**

**site** Unique study site ID

**AES\_type** AES habitat type

**lat\_AES** latitude coordinate (WGS84)

**lon\_AES** longitude coordinate (WGS84)

**area\_AES** area of AES habitats in ha (1 ha = 10000m²)

**percentage\_SNH** percentage of semi-natural habitats in 1km radius

**percentage\_arable** percentage of arable land in 1km radius

**elevation** elevation in m

1. **metabarcoding\_table:**

**phylum\_name** Phylum

**class\_name** Class

**order\_name** Order

**family\_name** Family

**BIN\_species** Individual taxa identifiers

**Columns 6 – 31** Individual study sites (numbers in columns represent reads in metabarcoding, not abundances)