

The zip file contains 5 csv data files of two categories. They contain either records list or taxa list.

Data files containing records lists:

- **RAINBIO.csv**

Contains all georeferenced records that passed the different quality checks of georeferencing.

NOTE: These records can be explored and exported in the following web application:

<https://gdauby.shinyapps.io/rainbio/>

- **RAINBIO_not_or_wrong_georef.csv**

Contains all records that did not pass the different quality checks of georeferencing.

- **Records_species_cultivated_introduced.csv**

Contains all records belonging to taxa tagged as introduced or cultivated.

Data files containing taxa lists:

- **species_list_table.csv**

Taxa list in RAINBIO.csv

- **liste_species_cultivated_introduced.csv**

List of taxa tagged as introduced or cultivated.

NOTE: all csv files are UTF8 encoded.

An **Rdata file** can be open in R and the script (**script_manip_database_published.R**) can be used to explore and export the database.

The zip file contains also a LOG csv (**log_rainbio20XX-XX-XX.csv**) file with the date of exporting. It contains several statistics of the database when it has been exported.

A pdf file (**Field explanations.pdf**) contains explanations for all field of the database.

Below are the meta-data with field explanations.

species list tables**dimensions: 10 columns, 29,656 entries**

column no.	name	comment
A	item number	unique row number for the file
B	tax_sp_level	genus and species not including infraspecific level
C	phylum	
D	order	
E	species	genus and species including infraspecific level
F	family	
G	genus	genus name. Split from column B
H	specificEpithet	species name. Split from column B
I	a_habit	controlled vocabulary: 'tree', 'shrub', 'shrublet', 'liana', 'epiphyte', 'herb', 'parasite', 'myco-heterotroph' consistent with field description of herbarium specimens (Dauby et al., 2016).
J	nbe_records	number of records held in the RAINBIO.csv

RAINBIO.csv / xlsx		
dimensions: 26 columns, 614,016 entries		
column no.	name	comment
A	item number	unique row number for the file
B	idrb	persistent unique identifier of the record in the RAINBIO database
C	phylum	
D	order	
E	family	
F	genus	
G	specificEpithet	
H	taxonRank	First rank of infra-specific level, with the following possible values: 'ESP': species level 'SUBSP': subspecies level 'VAR': variety level
I	tax_sp_level	genus and species not including infraspecific level
J	species	genus and species including infraspecific level
K	country	
L	countryCode	ISO 3166-1 alpha-3 country code indicated by the collection.
M	iso3lonlat	ISO 3166-1 alpha-3 country code derived from the geographical coordinates
N	decimalLatitude	The geographic latitude (in decimal degrees) of the locality. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.
O	decimalLongitude	The geographic longitude (in decimal degrees) of the record. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive
P	a_habit	controlled vocabulary: 'tree', 'shrub', 'shrublet', 'liana', 'epiphyte', 'herb', 'parasite', 'myco-heterotroph' consistent with field description of herbarium specimens and manually encoded by data providers (Dauby et al., 2016).
Q	a_habit_secondary	secondary life form of the species using classification above (see Dauby et al., 2016 for explanation)
R	catalogNumber	The unique identifier used in the source (child) datasets (idc). ID of the source database in capital letters + ID of the source database. 'JW' and 'JZ': Naturalis database 'TS': Tropicos database 'GD': Occurrences from tree plot data 'BR' and 'MS': Meise Botanic Garden database 'VD': Orchid database 'BS': Rubiaceae database 'DH': Dzanga-Sangha vascular plant database 'OH': Collection of African plant samples dried in silica-gel 'AO': African Palms 'BK': African Berlinia (Caesalpinioidae) 'UB': Université Libre de Bruxelles herbarium database (BRLU) 'KW': Royal Botanic Gardens, Kew (K) 'LISC': Instituto de Investigação Científica Tropical (LISC), University of Lisbon

S	institutionCode	<p>A character field indicating the source of the record :</p> <p>'Naturalis': Naturalis database</p> <p>'MO': Tropicos database (Missouri Botanical Garden)</p> <p>'Plots': Occurrences from tree plot data</p> <p>'BR': Meise Botanic Garden database</p> <p>'Orchids': Orchid database</p> <p>'Rubiaceae': Rubiaceae database</p> <p>'Dzanga': Dzanga-Sangha vascular plant database</p> <p>'OH_Silica': Collection of African plant samples dried in silica-gel</p> <p>'Palms': African Palms</p> <p>'Berlinia': African Berlinia (Caesalpinioidea)</p> <p>'BRLU': Université Libre de Bruxelles herbarium database (BRLU)</p> <p>'K': Royal Botanic Gardens, Kew (K)</p> <p>'LISC': Instituto de Investigação Científica Tropical (LISC), University of Lisbon</p>
T	idsc	If there are duplicates of the record, this field contains the list of other catalogue numbers (see catalogNumber)
U	calc_accuracy	<p>calculation accuracy of location from Table 1 of Dauby et al. (2016): to degree only (~110 km): 1; 15 minutes precision (~30 km): 2; 5 minutes precision (~10 km): 3; 2 minutes precision (~4 km): 4; minute is exact (~ 2 km): 5; 1/10 decimal minute exact (~ 200 m): 6; 1/100 decimal minute exact (~ 20 m): 7; 1/1000 decimal minute exact (~ 2 m): 8.</p>
V	morpho_names	<p>t': for TRUE, means the species is considered as a morpho species or not published species (i.e. such as <i>Deinbollia</i> sp. a, <i>Poggea</i> spec. nov.1, <i>Urelytrum debarbatum</i>, etc...); 'f': for FALSE.</p>
W	kind_col	<p>Controlled vocabulary of the specific nature of the data record: 'herb': herbarium collection; 'Observation'; 'unknown'; 'Sili': silica-gel sample ; 'Plot_data': data from plot.</p>
X	basisOfRecord	<p>Kind of record, following terminology of DarwinCore format: 'PreservedSpecimen', including 'herb' and 'Sili'; 'HumanObservation', including 'Plotdata' and 'Observation'</p>
Y	tax	Original submitted taxa name, superceded by verified names (cols C-J)
Z	fam	Original submitted family name, superceded by verified names (cols C-J)

RAINBIO not GEOREF.csv / xlsx**dimensions 28 columns, 157,885 entries**

column no.	name	comment most in common with RAINBIO)
A	item number	unique row number for the file
B	idrb	persistent unique identifier of the record in the RAINBIO database
C	phylum	
D	order	
E	family	
F	genus	
G	specificEpithet	
H	taxonRank	First rank of infra-specific level, with the following possible values: 'ESP': species level 'SUBSP': subspecies level 'VAR': variety level
I	tax_sp_level	genus and species not including infraspecific level
J	species	genus and species including infraspecific level
K	country	
L	countryCode	ISO 3166-1 alpha-3 country code indicated by the collection.
M	iso3lonlat	ISO 3166-1 alpha-3 country code derived from the geographical coordinates
N	decimalLatitude	The geographic latitude (in decimal degrees) of the locality. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.
O	decimalLongitude	The geographic longitude (in decimal degrees) of the record. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive
P	a_habit	controlled vocabulary: 'tree', 'shrub', 'shrublet', 'liana', 'epiphyte', 'herb', 'parasite', 'myco-heterotroph' consistent with field description of herbarium specimens and manually encoded by data providers (Dauby et al., 2016).
Q	a_habit_secondary	secondary life form of the species using classification above (see Dauby et al., 2016 for explanation)
R	catalogNumber	The unique identifier used in the source (child) datasets (idc). ID of the source database in capital letters + ID of the source database. 'JW' and 'JZ': Naturalis database 'TS': Tropicos database 'GD': Occurrences from tree plot data 'BR' and 'MS': Meise Botanic Garden database 'VD': Orchid database 'BS': Rubiaceae database 'DH': Dzanga-Sangha vascular plant database 'OH': Collection of African plant samples dried in silica-gel 'AO': African Palms 'BK': African Berlinia (Caesalpinioidea) 'UB': Université Libre de Bruxelles herbarium database (BRLU) 'KW': Royal Botanic Gardens, Kew (K) 'LISC': Instituto de Investigação Científica Tropical (LISC), University of Lisbon

S	institutionCode	<p>A character field indicating the source of the record : 'Naturalis': Naturalis database</p> <p>'MO': Tropicos database (Missouri Botanical Garden)</p> <p>'Plots': Occurrences from tree plot data</p> <p>'BR': Meise Botanic Garden database</p> <p>'Orchids': Orchid database</p> <p>'Rubiaceae': Rubiaceae database</p> <p>'Dzanga': Dzanga-Sangha vascular plant database</p> <p>'OH_Silica': Collection of African plant samples dried in silica-gel</p> <p>'Palms': African Palms</p> <p>'Berlinia': African Berlinia (Caesalpinioidea)</p> <p>'BRLU': Université Libre de Bruxelles herbarium database (BRLU)</p> <p>'K': Royal Botanic Gardens, Kew (K)</p> <p>'LISC': Instituto de Investigação Científica Tropical (LISC), University of Lisbon</p>
T	idsc	If there are duplicates of the record, this field contains the list of other catalogue numbers (see catalogNumber)
U	calc_accuracy	calculation accuracy of location from Table 1 of Dauby et al. (2016): to degree only (~110 km): 1; 15 minutes precision (~30 km): 2; 5 minutes precision (~10 km): 3; 2 minutes precision (~4 km): 4; minute is exact (~ 2 km): 5; 1/10 decimal minute exact (~ 200 m): 6; 1/100 decimal minute exact (~ 20 m): 7; 1/1000 decimal minute exact (~ 2 m): 8.
V	morpho_names	't': for TRUE, means the species is considered as a morpho species or not published species (i.e. such as <i>Deinbollia</i> sp. a, <i>Poggea</i> spec. nov.1, <i>Urelytrum debarbatum</i> , etc...); 'f': for FALSE.
W	kind_col	Controlled vocabulary of the specific nature of the data record: 'herb': herbarium collection; 'Observation'; 'unknown'; 'Sili': silica-gel sample ; 'Plot_data': data from plot.
X	basisOfRecord	Kind of record, following terminology of DarwinCore format: 'PreservedSpecimen', including 'herb' and 'Sili'; 'HumanObservation', including 'Plotdata' and 'Observation'
Y	tax	Original submitted taxa name, superceded by verified names (cols C-J)
Z	fam	Original submitted family name, superceded by verified names (cols C-J)
AA	verif_iso3	Result of georeferencing checking (whether or not the record falls within the recorded country), with the following possible values: 'OK': Coordinates fall in the same country indicated in the collection; 'NO COORDINATES'
AB	verif_coast	Result of georeferencing checking (whether or not the record fall within the ocean), with the following possible values: 'OK': Coordinates fall in the land; 'NO COORDINATES' ; 'ERROR': Coordinates fall in ocean (with continent with a buffer of 1 km).
AC	verif_distance	If 'verif_coast' or 'verif_iso3' indicate 'ERROR' or 'NEIGHBOUR', this field indicate the distance in meters to coast or border of the correct country.

liste_species_cultivated_introduced.csv		
dimensions	4 columns, 1566 items	
column no.	name	comment
A	tax_sp_level	genus and species not including infraspecific level
B	species	genus and species including infraspecific level
C	family	
D	a_habit	controlled vocabulary: 'tree', 'shrub', 'shrublet', 'liana', 'epiphyte', 'herb', 'parasite', 'myco-heterotroph' consistent with field description of herbarium specimens (Dauby et al., 2016).

Records_species_cultivated_introduced.csv		
dimensions	14 columns, 21,363 entries	
column no.	name	comment
A	idrb	persistent unique identifier of the record in the RAINBIO database
B	decimalLatitude	The geographic latitude (in decimal degrees) of the locality. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.
C	decimalLongitude	The geographic longitude (in decimal degrees) of the record. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive
D	tax_sp_level	genus and species not including infraspecific level
E	family	
F	genus	
G	specificEpithet	
H	species	genus and species including infraspecific level
I	country	
J	countryCode	ISO 3166-1 alpha-3 country code indicated by the collection.
K	verif_iso3	Result of georeferencing checking (whether or not the record falls within the recorded country), with the following possible values: 'OK': Coordinates fall in the same country indicated in the collection; 'NO COORDINATES'
L	verif_coast	Result of georeferencing checking (whether or not the record fall within the ocean), with the following possible values: 'OK': Coordinates fall in the land; 'NO COORDINATES' ; 'ERROR': Coordinates fall in ocean (with continent with a buffer of 1 km).
M	verif_distance	If 'verif_coast' or 'verif_iso3' indicate 'ERROR' or 'NEIGHBOUR', this field indicate the distance in meters to coast or border of the correct country.
N	locality	locality where the specimen has been collected