

# Data mobilisation from GBIF to the EBV Data Portal

Harmonisation of species keys for the JSON query of species listed in the Annex I of the Birds Directive

true

2024-08-23

## Introduction

In this notebook we prepare the **speciesUsageKey** for birds listed in Annex I of the Birds Directive. Subsequently, we will work with the GBIF occurrence data and therefore use the **speciesUsageKey** in its backbone taxonomy. We have previously prepared a list from Annex I of the Birds Directive that includes the taxonomy.

*Note: This series of notebooks is part of the results of Task 3.3 of the Biodiversity Building Blocks for Policy project funded by the European Union's Horizon Europe Research and Innovation Programme (ID No 101059592). Additional notebooks exploring the results and calculating simple metrics are also available in the same repository.*

## Load Libraries

```
rm(list=ls())
gc()
```

```
##           used  (Mb) gc trigger  (Mb) max used  (Mb)
## Ncells 2270313 121.3   6147489 328.4 12006812 641.3
## Vcells 5136122  39.2   43069612 328.6 104364218 796.3
```

```
# Load requiered libraries
library(here)
library(dplyr)
```

## Data Pre-processing

First we start loading the species taxonomy obtained through the gbif species matching tool.

```
# Load species taxonomy only with 'accepted' scientific names in the GBIF backbone taxonomy
tax <- read.csv(here("input/data/birds/taxonomy/list193birds_directive_annexi_match_gbif_updated.csv"))
```

As some of the scientific names in Annex I of the Birds Directive are considered synonyms by the GBIF backbone taxonomy, we will use the `acceptedUsageKey` for the synonyms and the `key` for the accepted scientific names. To do this, we will fill in the `acceptedUsageKey` column with the key for accepted names, and keep the `acceptedUsageKey` for synonyms. Thus, the accepted keys for all species in the list will appear in the `acceptedUsageKey` column.

```
# Merge in one column `key` of accepted names and `acceptedUsageKey` of synonyms
usagekeys <- tax %>%
  mutate(acceptedUsageKey = coalesce(acceptedUsageKey, key))

# Write CSV with `acceptedUsageKey` for all species to be used in the JSON query
write.csv(usagekeys, here("input/data/birds/taxonomy/list193birds_directive_annexi_allaccepted_usagekey."
```