The files contained in this folder contains the Supplementary Information for the study entitled “COVID-19 has led to a global increase in web searches for bats: a risk for conservation ?”, archived on EcoEvoRxiv (<https://doi.org/10.32942/osf.io/4pa2q>).

Files include**:**

* a table of media coverage in Italy, about the enter into force of the first Union List of invasive alien species (n.1143/2016) (“MediaCoverage.csv”). The table reports various articles that were published by Italian media between July and September 2016, following the Union List. For each article we provided a link to the original source.
* the reproducible script for the statistical software R, to replicate our analyses. The script requires the installation of the statistical software R, as well as INLA (<https://www.r-inla.org/>) and various other packages, that are reported in the code.
* 4 datasets that can be uploaded during some section of data analysis. From the R script it will be possible to download raw data. However, we set up some datasets for quick data analysis, as downloading data from Google, Wikipedia and GDELT requires access to the Internet. These datasets are:
  + a dataset about overall news coverage about bats, among US newscasts, downloaded from GDELT (GDELTBatsUS1620.csv). The dataset is necessary to produce the plot shown in Fig. 1 of the preprint. The dataset contains four different fields:
    - * network: which networks were queried by GDELT. In this case all the available ones (the only value is “Combined”)
      * date: the date of each value of the time series
      * value: the value of the GDELT index
  + a dataset containing the GoogleTrends time series, for the various queries that were searched in the 20 different countries of our study (“GoogleTrends1620\_Bats.csv”). The dataset was produced with the R package ‘gtrendsR’: <https://rstudio-pubs-static.s3.amazonaws.com/155168_61e1f687681d44e4988ce28b7f6ec13b.html>
  + The daset contained the following variables:
    - * date: the date of each point of the time series
      * hits: the value of the GoogleTrends index
      * keyword: the query that was searched for on Google
      * geo: the country for which the GoogleTrends index of a certain query was extracted
      * time: the timespan of our query (min, max)
      * gprop: the source of data extraction (the web, in this case, corresponding to the various extensions of Google)
  + a dataset containing values about visits to the various Wikipedia pages about bats, reported in in Table 1 (“WikipediaWeekly1620\_Bats.csv”). The dataset was extracted from PageViews and SiteViews (<https://pageviews.toolforge.org/>), and it was assembled by aggregating daily observations into weekly sums of visits. The dataset contains a series of variables, in the wide format:
    - * date: the date of each point of the time series
      * tot.fr: the total number of visits to the French Wikipedia
      * bats.fr: the total number of visits to the page about bats on the French Wikipedia (the number of visits to Wikipedia pages about bats, per 1,000,000 total visits can be obtained by dividing these two columns and then multiplying them per 1,000,000)
      * tot.en / bats.en: the same for English Wikipedia
      * tot.de / bats.de: the same for German Wikipedia
      * tot.es / bats.es: the same for Spanish Wikipedia
      * tot.it / bats.it: the same for Italian Wikipedia
      * tot.pt / bats.pt: the same for Portuguese Wikipedia
      * tot.ja / bats.ja: the same for Japanese Wikipedia
      * tot.ko / bats.ko: the same for Korean Wikipedia
  + a dataset containing the Topic of the relates searches extracted from Google, for the queries at Table 1. The dataset was extracted with ‘gtrendsR’, see above and it contained the following variables:
    - keep: a variable indicating whether a certain observation should have been removed (0) or kept (1). We kept only topics that were assigned a GoogleTrends score (field “subject”)
    - subject: the value of the GoogleTrends index, indicating the popularity of the various topics
    - related\_topics: a variable indicating whether popularity was evaluated on a stable basis or not
    - value: a variable indicating the topic of related searches, from the most common, to the less common
    - geo: a variable indicating the country from which data were extracted (country is the full name, geo the acronym)
    - year: a variable indicating the year for which the various topics were extracted and compared in their popularity
  + a dataset containing the keywords characterizing news about “bats” that were extracted from GDELT. Data were extracted with the newsflash package: <https://github.com/hrbrmstr/newsflash>. The dataset included the following variables:
    - label: a variable indicating the single keyword
    - count: a variable indicating how the single keyword was popular within each years
    - year: the year for which keywords were extracted and compared