SPI-Birds Newsletter

Welcome to the October edition of the SPI-Birds newsletter. With the weather getting colder, we've had more time inside to work on the SPI-Birds project. Over the last two months our database has expanded with the addition of new data, pipelines, and meta-data. We've also been working to improve our data quality check to produce a clear report for data owners. Read more about our progress below.

Newest pipelines

In the last months we have added new pipelines for the populations at CEFE, Montpellier, which includes the well known populations in Corsica and Rouviere. We've also included data from within the Arctic circle, adding a pipeline for Kevo, Finland. With the inclusion of data from Montpellier and Kevo we now have pipelines for a total of 21 populations, covering 9 countries ranging from the Mediterranean Sea (37°36' N) to the Arctic circle (69°45' N). New data is still coming in, and we will be adding at least 2 more countries to this list before long. Keep updated on Twitter and the SPI-Bird website.

Pipeline code available online

Pipelines for all 21 populations are now available online at the SPI-Birds GitHub page. Have a look at our GitHub and explore the pipelines! To be clear, this is **only** the pipeline code, and does not include any primary data. The introduction included on the GitHub page gives you an idea how to load the pipelines on your computer and understand how the data have been processed. This includes preparing your computer for running the SPI-Birds pipelines, loading the pipelines package in R and accessing help files for each pipeline.

Quality checks

In addition to the pipelines, the GitHub page also includes the code for our data quality check process. Attached to the e-mail with this newsletter is an example of the quality check report carried out on data from Hoge Veluwe in the years XXX - YYY. We have separated the quality check to include 'warnings' (values that we believe are unlikely) and 'potential errors' (values that we believe are impossible e.g. negative values). The quality check report also includes a brief summary of the types of checks carried out,

which have been based on the checks carried out by the Evolutionary Ecology Group at the University of Antwerp. We are still hoping to add additional checks. If you have any ideas, please contact us.

The quality check report will serve two purposes. Firstly, we will send a copy of the quality check report to any user that requests data. This will allow data users to easily identify and (if necessary) remove any potentially spurious records. We will also send versions of this quality check report to data owners, which we hope will help improve the quality and integrity of primary data. These quality check reports are intended to provide extra information to data owners, and there is no requirement for these errors and warnings to be addressed; however, if the report helps you identify potential mistakes in the data, please keep a record of any changes that were made. Before the beginning of the next breeding season we will send a request to all data owners for an updated version of the primary data. At this time, please let us know of any changes made to the primary data stemming from the quality check report. Additionally, if you notice any records flagged as 'potential errors' or 'warnings' that are true observations, please also highlight these so that they are not flagged again in future quality check reports.

SPI-Birds paper

We are now in the process of writing a paper about the SPI-Birds project. In the paper, we will use SPI-Birds as an example of good practice in open science and scientific reproducibility, with specific information about how we approached the SPI-Birds project and our workflow for integrating multiple different primary datasets into a single standard format. We want this to be a community project with contributions from the members of the SPI-Birds network. In the coming months, we will create a draft version of the manuscript (likely using Google Docs) and we encourage you to contribute your thoughts.

How can you contribute?

If you're interested in contributing towards the SPI-Birds project, here's what you can do:

- Provide us with access to your primary data from which we can build a pipeline. This pipeline will convert your primary data into the *Standard protocol* described above. To be clear, these data will not be shared, distributed, or used for any purpose without your direct permission. If you still have any questions or concerns about the process, please contact our co-ordinator Antica Culina (A.Culina@nioo.knaw.nl).

- If you have already provided access to your primary data, tell us some interesting facts or describe a seminal paper that involves your population. We will use this information to introduce your population on Twitter.

- Follow us on Twitter.

- Let us know of any papers, conferences, or workshops that you feel will be interesting to SPI-Birds members.

Thank you again for your contribution and support so far. We look forward to sending you another update in two months time.