

APPROVED: 15 September 2023  
doi: 10.2903/sp.efsa.20YY.EN-NNNN

# *shinyIVT*: Development of a webApp to validate ENETWILD data

Sergio López-Padilla<sup>1</sup>, José Antonio Blanco-Aguilar<sup>1</sup>, Pelayo Acevedo<sup>1</sup>, Joaquín Vicente<sup>1</sup>.

<sup>1</sup>IREC- National Institute of Game and Wildlife Research

## Abstract

The ENETWILD project aims to create an integrated platform for collecting, managing, and sharing population data on wildlife across Europe. As part of this effort, the project has developed the Integrated Validation Tool (IVT), a web application that enables researchers and data managers to validate data collected for the project according to Darwin Core guidelines and standards. The IVT provides a comprehensive and user-friendly solution for automated validation checks for metadata and data files. The IVT's modular design and comprehensive validation checks ensure that users can submit high-quality data that adheres to Darwin Core guidelines. Its spatial data validation checks generates an output ZIP file with corrected data, and a validation report to enhance its usability and efficiency. As the IVT continues to evolve, it is recommended the development of additional functionalities, such as a spatial data visualization tab and the integration with other data sources, as well as incorporating user feedback for future updates.

© European Food Safety Authority, 2023

**Key words:** Enetwild, webApp, European wildlife, data validation, on-going development, WildnetR.

**Question number:** EFSA-Q-YYYY-NNNNN

**Correspondence:** xxx@efsa.europa.eu

**Disclaimer:** The present document has been produced and adopted by the bodies identified above as author(s). This task has been carried out exclusively by the author(s) in the context of a contract between the European Food Safety Authority and the author(s), awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

**Acknowledgements:**

**Suggested citation:** López-Padilla S, Blanco-Aguilar JA, Acevedo P, Vicente J, 2023. shinyIVT: Development of a webApp to validate ENETWILD data. EFSA supporting publication 20YY:EN-NNNN. 15 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN

**ISSN:** 2397-8325

© European Food Safety Authority, 2023




## Summary

The *ENETWILD* project aims to create an integrated platform for collecting, managing, and sharing population data on wildlife across Europe. As part of this effort, the project has developed the Integrated Validation Tool (IVT), a web application that enables researchers and data managers to validate data collected for the project according to Darwin Core guidelines, i.e. data format are following the data structure proposed by ENETWILD for Wildlife data model (ENETWILD consortium et al 2020 ). The app is presented in this report. The IVT is a user-friendly web application that streamlines the validation process by providing automated checks for metadata and data files. The IVT's workflow is designed to guide users through the validation process in a step-by-step manner, and it includes several modules for uploading and validating data. The IVT leverages a range of R packages to enable data validation, including dplyr, sf, shinyjs, rgdal, uuid, readr, and assertr, among others.

*ENETWILD* is committed to ensuring that users have the tools they need to collect and manage high-quality data on wildlife across Europe. To achieve this goal, the project has developed a template for data collection that will ensure standardization and consistency in the data collected by users. The template consists of five sheets: Metadata, draftEW (data), Filling tools, vocabulary metadata, and vocabulary draftEW. The metadata sheet contains information about the data collection methodology and other relevant details on the data. The draftEW sheet contain the actual data collected harmonized by the data providers. The filling tools sheet provide guidance on how to fill out the data sheet, including examples of correct formatting. The vocabulary metadata and vocabulary data sheets will provide users with a list of allowed values for certain fields in the metadata and data sheets, respectively, to ensure consistency and standardization.

The IVT represents a significant step forward in ENETWILD's efforts to provide an integrated platform for collecting, managing, and sharing data on wildlife across Europe. The IVT's modular design, as well as comprehensive validation checks, ensure that users can submit high-quality data that adheres to Darwin Core guidelines. Its spatial data validation checks generate an output ZIP file with corrected data, and a validation report to enhance its usability and efficiency. This first version of the IVT will continue to be updated and refined based on user feedback and testing. The ongoing development of the IVT will focus on improving its speed, efficiency, and user-friendliness, as well as expanding its functionality to accommodate additional types of data and validation checks. To make the IVT more accessible to users, it is recommended to develop a comprehensive user manual, as well as a video tutorial to illustrate how to fill out and validate the data in the Excel template. In future versions of the IVT, it is also recommended that a spatial data visualization tab be added to allow users to



visualize their data and identify potential issues. This could include a distribution map and other interactive visualizations that allow users to explore their data in greater detail. To ensure that the IVT continues to meet the needs of its users, it is recommended to ask user feedback regularly, and to incorporate it into future updates. The IVT will be subject to thorough testing and validation to ensure that its functions are working as intended and providing accurate results.

## Table of contents

<i>Abstract</i> .....	1
<i>Summary</i> .....	3
<i>1. Introduction</i> .....	6
1.1. Background and terms of reference as provided by the requestor.....	6
2.1. Data Model template (WLDM_1.5.2).....	7
2.2. Methods.....	8
<i>3. Assessment/Results</i> .....	9
<i>4. Conclusions</i> .....	13
<i>5. Recommendations</i> .....	14
<i>References</i> .....	15

## 1. Introduction

*ENETWILD* is a European research project that seeks to create a platform for the collection, management, and sharing of data on wildlife populations across Europe ([www.enetwild.com](http://www.enetwild.com)). To achieve this goal, *ENETWILD* is developing several web applications, including an interactive map for density and abundance data on different species. This map consists of two modules: one for authorized EFSA staff and another publicly accessible module for users to consult available data without confidentiality issues. To ensure data accessibility, the *ENETWILD* team developed *WildenetR*, an R package that allows users to download data in different formats and perform advanced data processing and analysis (López-Padilla et al. 2023).

As part of these initiatives, *ENETWILD* has developed in this report the Integrated Validation Tool (IVT), a web app that enables data providers (users) to validate data at submission according to the proposed ENETWILD Data model (ENETWILD consortium et al 2020). IVT consists of several modules that allow users to upload and validate data, including metadata and spatial data. Once the validation checks are completed, users receive a .zip file with the validated data and a text file with additional information for the coordinating data managers on how the data was validated or if there are any warning to be assessed, as a preliminary step before incorporation into the main database. By developing these web applications and tools, *ENETWILD* aims to provide an integrated platform for collecting, managing, and sharing data on wildlife across Europe. The IVT, along with other developments, like *WildenetR* (an R package that allows users to download and process data from the *Enetwild* database) and the interactive map, is a crucial component of this platform, providing researchers and data managers with powerful tools to ensure data quality and accessibility.

### 1.1. Background and terms of reference as provided by the requestor

This contract was awarded by EFSA to Universidad de Castilla-La Mancha, contract title: Wildlife: collecting and sharing data on wildlife populations, transmitting animal disease agents, contract number: OC/EFSA/ALPHA/2016/01 – 01. The terms of reference for the present report (specific contract 11, deliverable 5.3) indicates “to develop a Web validation tool (e.g., as R shiny app) to facilitate data submission and validation. This tool validates the data uploaded by the data providers (e.g., IPT from GBIF). This tool provides feedback (report) when the uploaded files do not comply with the established model, or if a warning is identified. If and when the data are uploaded in the correct format, the data can be transferred to the temporal database”.

## 2. Data and Methodologies

### 2.1. Data Model template (WLDM\_1.5.2)

*ENETWILD* is committed to ensuring that users have the practical tools they need to collect and manage high-quality data on wildlife across Europe. As part of this effort, the project is developing a template for data collection that will ensure standardization and consistency in the data collected by users.

The template will consist of five sheets: Metadata, draftEW (data), Filling tools, vocabulary metadata, and vocabulary draftEW. The metadata sheet will contain information about the data collection methodology and other relevant details of the data. The draftEW sheet will contain the actual data collected harmonized by the data providers. The filling tools sheet will provide guidance on how to fill out the data sheet, including examples of correct formatting. The vocabulary metadata and vocabulary data sheets will provide users with a list of allowed values for certain fields in the metadata and data sheets, respectively, to ensure consistency and standardization.

By providing a standardized data template, *ENETWILD* aims to simplify the data collection process and ensure that data collected across the project is of the highest quality in term of standardization and harmonisation. The template will be designed to be user-friendly and accessible, with clear guidance and instructions for users. The project team is working diligently to finalize the template so that it can be made available to users as soon as possible

## 2.2. Methods

The development of the Integrated Validation Tool (IVT) relied on a range of tools and methodologies to ensure that the final product was robust, reliable, and user-friendly. The use of Docker allowed us to create a consistent and reproducible development environment, while the adoption of Git as a version control system enabled collaboration between developers and ensured that changes to the codebase were properly tracked and documented.

The development process followed best practices for R package development, including the use of modular programming to facilitate code reuse, and the adoption of the Shiny framework to enable the development of a responsive and interactive user interface. The IVT leverages a range of R packages, including dplyr, sf, shinyjs, rgdal, uuid, readr, and assertr, among others, to enable data validation according to Darwin Core guidelines (ENETWILD-consortium et al. 2020, 2022).

To ensure the quality and reliability of the IVT, we implemented a range of testing and validation procedures throughout the development process. These included unit testing of individual functions, integration testing of different modules, and user acceptance testing of the final product.

Overall, the development of the IVT relied on a range of methodologies and tools to ensure that the final product was reliable, user-friendly, and fully aligned with the requirements of the ENETWILD project.



### 3. ASSESSMENT/RESULTS

The IVT developed by ENETWILD offers a comprehensive and user-friendly solution for researchers and data managers to validate collected data according to Darwin Core guidelines. The application is designed to streamline the validation process by providing a user-friendly interface and automated validation checks for metadata and data files. The repository with all the files is available at the IREC-CSIC-UCLM repository: <https://github.com/IREC-CSIC-UCLM/shinyIVT>.

The IVT comprises of several directories, modules, and scripts, which work together to enable users to submit, validate, and download their data. The IVT's data directory contains several files, including the `draftEW_dict.rds`, `metadata_dict.rds`, and `WLDM_1.5.2.xlsx` (ENETWILD-consortium et al. 2022). These files provide the application with necessary information and parameters to validate the data files. The modules directory contains server and user-interface UI scripts that provide the functional components of the application. These scripts are used to create the UI and handle the submission, validation, and download processes.

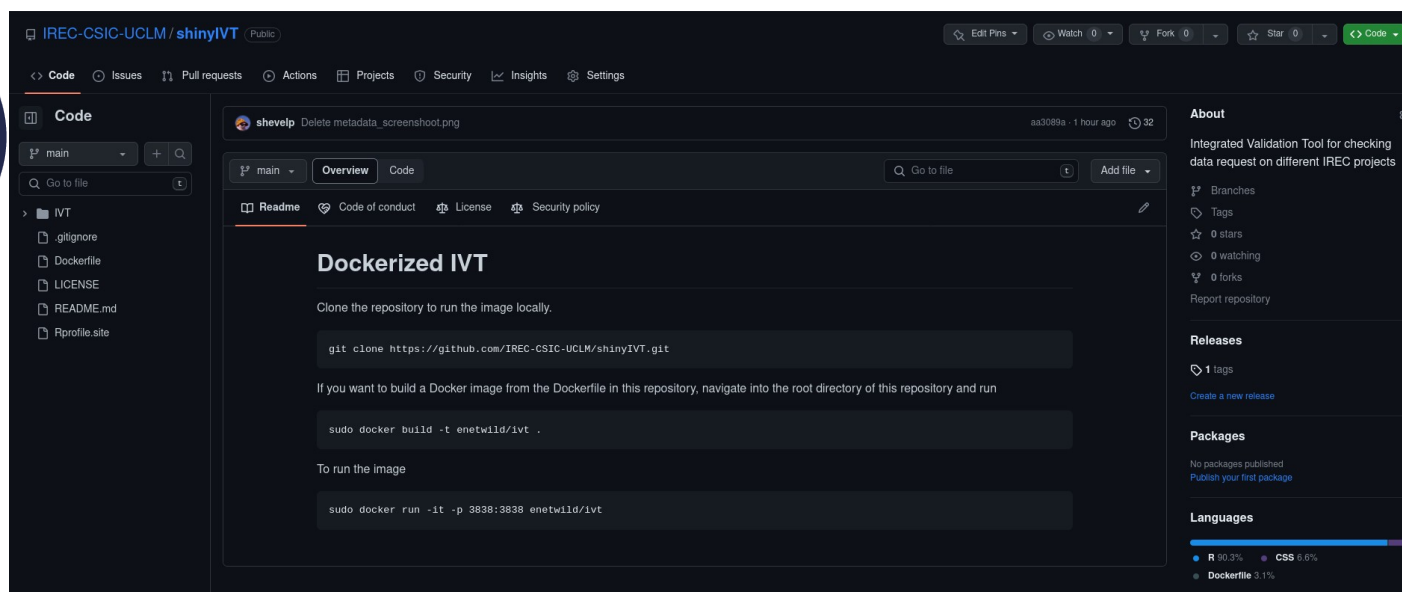
The IVT's server directory includes server-side scripts for validating data, including the `dataValidateServer.R`, `submitPanelDynamicServer.R`, `metadataValidateServer.R`, `submitPanelValidateServer.R`, and `templateIRECAppsServer.R` scripts. These scripts enable the application to handle the validation of metadata and data files, as well as provide users with detailed feedback on the validation checks.

The IVT's UI directory includes the `submissionPanelUI.R` and `templateIRECAppsUI.R` scripts that handle the user interface components of the application. The `submissionPanelUI.R` script creates the submission panel where users can enter their data, while the `templateIRECAppsUI.R` script provides the application's main interface.

The R directory contains several scripts, including the `data_processing.R`, `validation.R`, and `utils.R` scripts. These scripts provide essential functionality for the IVT's server-side validation, data processing, and data download functionality.

The IVT's tabs directory includes server-side and UI scripts that provide the main interface of the application. These scripts include the `server_IVT.R` and `ui_IVT.R` scripts.

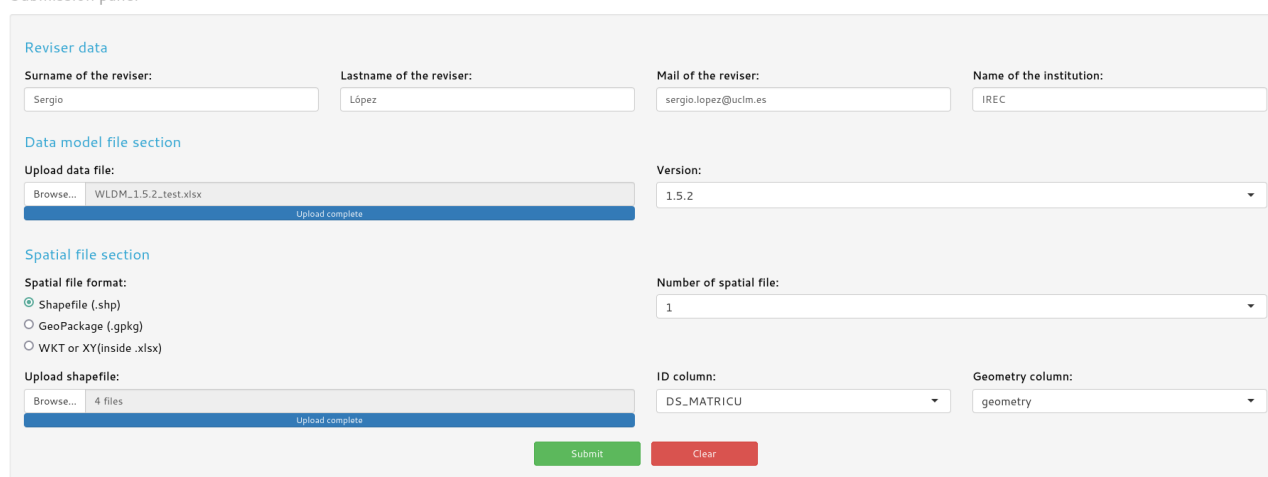
The IVT's www directory includes several static files used in the application, including the `manual.pdf`, `styles.css`, and `txt` files.



**Figure 1:** Code repository of the IVT webApp (screenshot).

The IVT application workflow is designed to guide users through the data validation process in a step-by-step manner. The user begins by filling out the submission panel, which contains fields for general information about the data being submitted, such as the name of the project, data provider, and data file(s). Once the user has filled out the submission panel, they can submit their data by clicking the submit button.

## Submission panel



The screenshot shows the 'Submission panel' of the webApp. It is divided into two main sections: 'Reviser data' and 'Data model file section'.

**Reviser data:**

- Surname of the reviser: Sergio
- Lastname of the reviser: López
- Mail of the reviser: sergio.lopez@uclm.es
- Name of the institution: IREC

**Data model file section:**

**Upload data file:**

- Browse...: WLDLM\_1.5.2\_test.xlsx
- Version: 1.5.2

**Spatial file section:**

**Spatial file format:**

- ☒ Shapefile (.shp)
- ☐ GeoPackage (.gpkg)
- ☐ WKT or XY (inside .xlsx)

**Upload shapefile:**

- Browse...: 4 files

**Number of spatial file:** 1

**ID column:** DS\_MATRICU

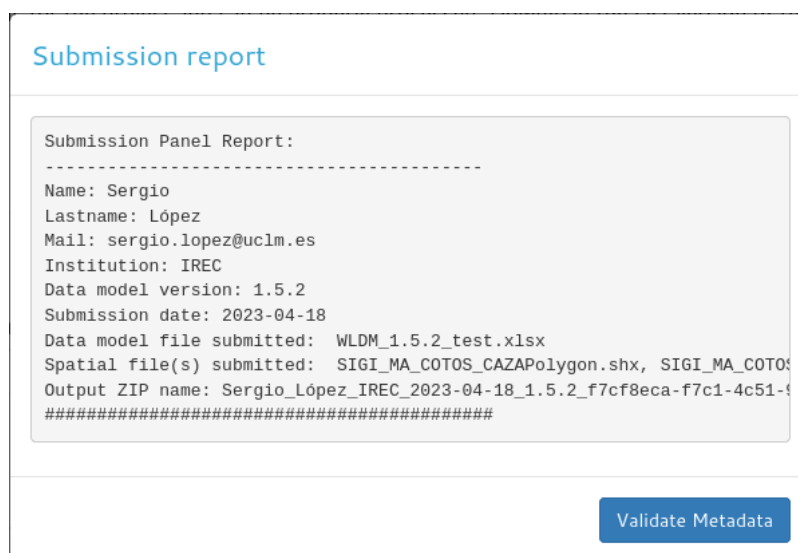
**Geometry column:** geometry

Buttons: Submit, Clear

© 2023 IVT | Twitter | Contact Us

**Figure 2:** Submission Panel of the webApp.

If the submission panel is filled out correctly, the user will see a dropdown summary of their submission and a button to validate the metadata. If there are errors in the submission panel, the user will receive an error message and will not be able to proceed until the errors are corrected.



The screenshot shows the 'Submission report' pop-up. It contains a summary of the submission data and a button to validate the metadata.

**Submission report**

Submission Panel Report:

```

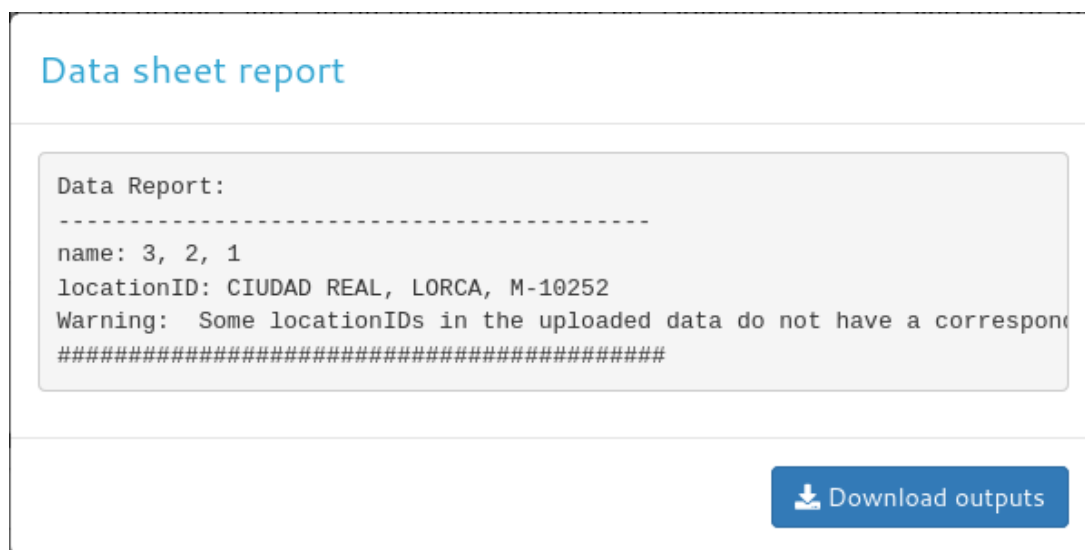
-----
Name: Sergio
Lastname: López
Mail: sergio.lopez@uclm.es
Institution: IREC
Data model version: 1.5.2
Submission date: 2023-04-18
Data model file submitted: WLDLM_1.5.2_test.xlsx
Spatial file(s) submitted: SIGI_MA_C0T0S_CAZAPolygon.shx, SIGI_MA_C0T0S_CAZAPolygon.shx
Output ZIP name: Sergio_López_IREC_2023-04-18_1.5.2_f7cf8eca-f7c1-4c51-9a31-3d8a1b1b1b1b
#####
  
```

Validate Metadata

**Figure 3:** Submission Panel pop-up after validation.

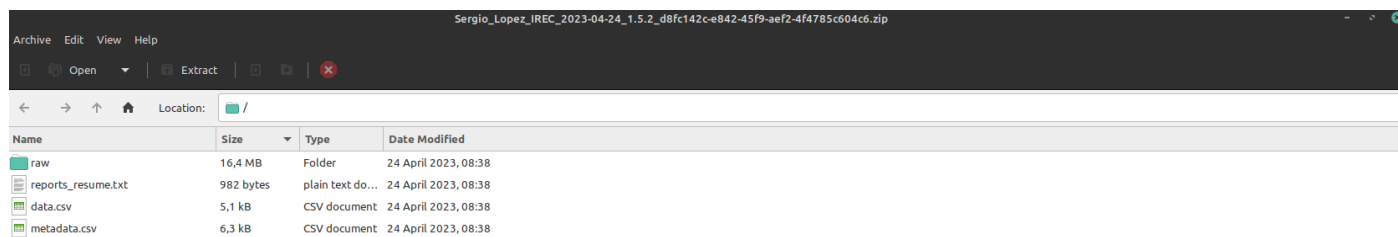
When the user clicks the "validate metadata" button, the application performs a series of validation checks, including structure, required values, and allowed values. If there are errors in the metadata, the user will receive an error message and will not be able to proceed until the errors are corrected. If the metadata passes validation, the user will see a summary of their metadata and will be prompted to validate their data.

If the user has uploaded a shapefile or geopackage, the application will perform a spatial union based on the selected column in the submission panel and will perform validation checks on the structure, required values, and allowed values of the data. If there are errors in the data, the user will receive an error message and will not be able to proceed until the errors are corrected. If the data passes validation, the user can download a .zip file containing the validated data, metadata, along with a summary of any warnings or errors encountered during the validation process and the raw files submitted by the user.



**Figure 4:** Data pop-up after validation.

The IVT application is designed to help researchers and data managers to validate their data in a comprehensive and user-friendly manner. By providing a structured and guided workflow, the IVT application ensures that data are validated according to best practices and can be easily shared and used by others.



**Figure 5:** Generated file in .zip format to upload to data coordinators for final checks and merging into the database.

## 4. Conclusions

- The Integrated Validation Tool (IVT) represents a significant step forward in *ENETWILD*'s efforts to provide an integrated platform for collecting, managing, and sharing data on wildlife across Europe.
- The IVT's modular design and comprehensive validation checks ensure that users can submit high-quality data that adheres to Darwin Core guidelines. The inclusion of spatial data validation checks, along with the ability to generate an output ZIP file with corrected data and a text file summarizing validation reports, greatly enhances the usability and efficiency of the tool.
- The first version of the IVT will continue to be updated and refined based on user feedback and testing. The ongoing development of the IVT will focus on improving its speed, efficiency, and user-friendliness, as well as expanding its functionality to accommodate additional types of data and validation checks.
- Overall, the IVT is an essential tool for researchers and data managers who need to ensure that their data adheres to Darwin Core guidelines and is of the highest possible quality. By providing a comprehensive and user-friendly validation process, the IVT enables users to confidently submit their data to *ENETWILD*'s database.

## 5. Recommendations

- The Integrated Validation Tool (IVT) is a valuable resource for researchers and practitioners able to provide data on wildlife across Europe. As the IVT continues to evolve, it is recommended that additional functionalities be developed to allow for even more comprehensive analysis of data. This could include the integration with other data sources, such as remote sensing or citizen science data, to provide more complete coverage of species distribution and abundance.
- To make the IVT more accessible to users, it is recommended that a comprehensive user manual be developed, as well as a video tutorial to illustrate how to fill out and validate the data in the Excel template. This will make it easier for new users to understand the submission process and ensure that they are providing accurate and complete data.
- In future versions of the IVT, it is recommended that a spatial data visualization tab be added to allow users to visualize their data and identify potential issues. This could include a distribution map and other interactive visualizations that allow users to explore their data in greater detail.
- To ensure that the IVT continues to meet the needs of its users, it is recommended to collect their feedback regularly, and to incorporate it into future updates. This feedback could be collected through user surveys or through the Issues section of the IVT's GitHub repository (<https://github.com/IREC-CSIC-UCLM/shinyIVT/issues>).
- Finally, it is recommended that the IVT be subject to thorough pilot testing and validation to ensure that its functions are working as intended and providing accurate results. This could include collaboration with other researchers and practitioners providing data on wildlife across Europe to identify potential issues and areas for improvement.

## References

*ENETWILD* consortium, Guillaume Body, Mathilde Mousset, Emmanuelle Chevallier, Massimo Scandura, Sophie Pamerlon, Jose Antonio Blanco-Aguilar, Joaquin Vicente, 2020. Applying the Darwin core standard to the monitoring of wildlife species, their management and estimated records. EFSA supporting publication 2020:EN-1841. 81 pp. doi:10.2903/sp.efsa.2020.EN1841.

*ENETWILD* consortium, Francesca Jaroszynska, Guillaume Body, Sophie Pamerlon, Anne-Sophie Archambeau, 2022. Applying the Darwin Core data standard to wildlife disease -advancements toward a new data model. EFSA supporting publication 2022:EN-7667. 68 pp. doi:10.2903/sp.efsa.2022.EN-7667.

López-Padilla S, Blanco-Aguilar JA, Acevedo P, Vicente J, 2023. WildnetR: Development of a R package to access *ENETWILD* data. EFSA supporting publication 20YY:EN-NNNN. 9 pp. doi:10.2903/sp.efsa.20YY.EN-NNNN.