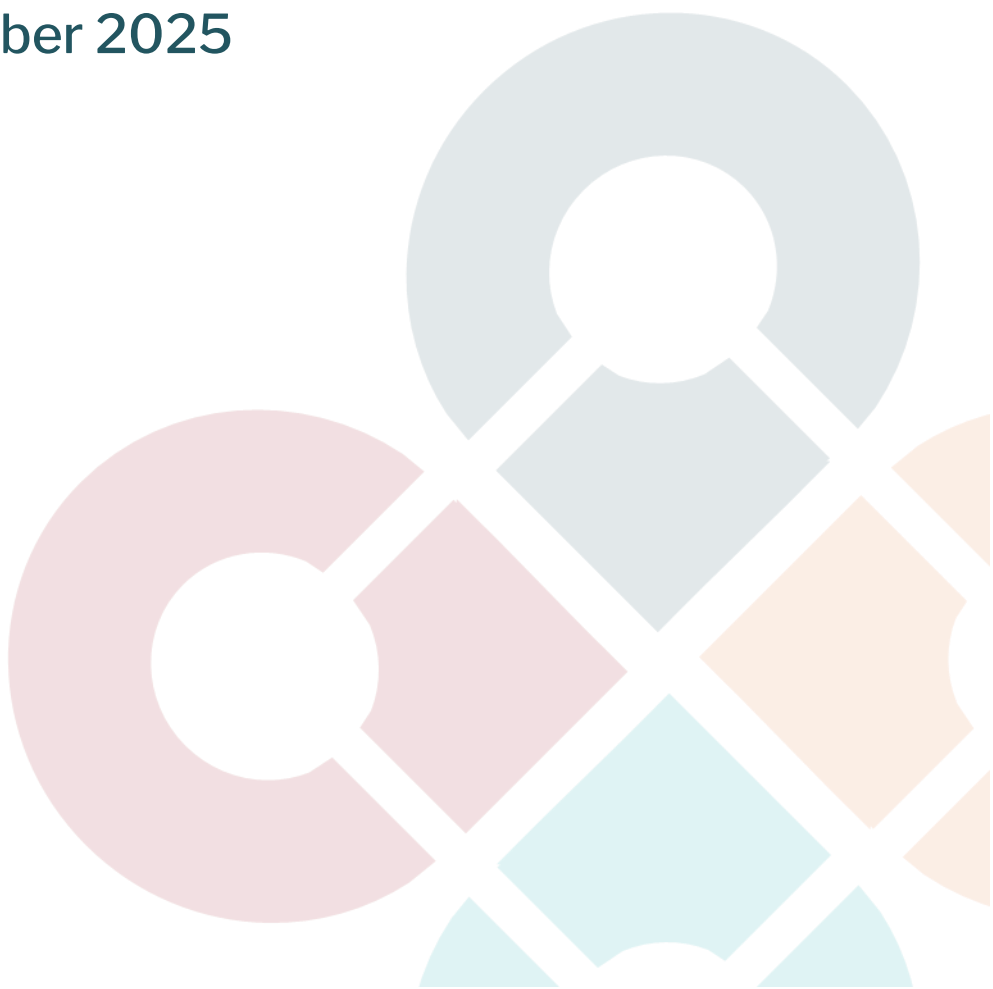# User Documentation

Vector Data Validation Tool

EOI/EFSA/2022/01 – CT 42 BIOHAW

Prepared for EFSA

By EpiMundi

1 December 2025

1 About EpiMundi

EpiMundi is an independent epidemiology research and consulting company, registered in France.

It was established as **Ausvet Europe**, part of the Ausvet Group in 2019. EpiMundi is now owned and led by its President, Dr Angus Cameron, who was the first partner at the founding of Ausvet in Australia in 1998.

Ausvet Europe was renamed as EpiMundi in August 2024 to better reflect the work of the company – providing epidemiological expertise across animal, human and plant health around the globe.

Formerly operating as

1.1 Our mission

In response to the growing challenges facing the planet, we are driven to make a significant positive impact on the health and welfare of animals and humans throughout the world by providing inspired, pragmatic, people-centred and evidence-based expertise in epidemiology.

1.2 Disclaimer

Except where explicitly specified, the contents of this document are confidential and are intended for the use of the client only. If you have received a copy of this report in error, please delete it and notify the sender. EpiMundi is an independent consulting company and research provider and is not aware of any conflict of interest in relation to the content of this document. The information provided is based on internal research, peer reviewed publications, and the broad experience of the EpiMundi team. While all professional care is taken, given the inherently unpredictable nature of biological systems and epidemiological research, no warranty is provided as to the accuracy of specific information. Any recommendations or findings that may be included in this document are based on available evidence and experience. EpiMundi provides no guarantee that actions taken on the basis of these recommendations and findings will result in the desired outcome in any specific case.

© 2025 EpiMundi

Contents

1 Introduction

Welcome to the VectorNet Data Validation System. This system is designed to help ensure that all vector surveillance data is accurate, consistent, and complete before it is submitted to the VectorNet GBIF repository.

It works by automatically checking your data spreadsheet for common errors and providing detailed feedback for correction. This guide explains how to use the system based on your role.

2 Guide for Data Providers (Entomologists)

This section is for entomologists and other data providers who collect data and submit it for validation.

2.1 Overview of Your Role

Your role in this process is to:

1. **Enter** your surveillance data into the standard Excel template.

2. **Submit** the template to the validation system by email.

3. **Correct** any errors identified by the system and resubmit.

4. **Approve** the validated data (with zero errors) for final submission to a Reference Entomologist.

This validation loop (Submit -> Correct -> Resubmit) can be repeated as many times as necessary.

2.2 Understanding the Excel Template

The Excel workbook is your primary tool for data entry. It contains built-in features to help you enter data correctly.

- **Worksheets:** The template contains several worksheets:

  - **0. Instruction:** Read this sheet first. It contains important help and instructions.

  - **1.DATA INPUT-Ticks:** This (or a similar sheet for your vector group) is the *only* sheet you need to edit with your data.

  - **2.NUTS3, 3.Vector species, 4.Host species:** These are customised reference lists used by the template. Do not edit them unless you are adding a new species (see below). A further Codes worksheet is hidden to avoid changes.

- **Built-in Help:**

  - **Header Information:** In the 1.DATA INPUT-Ticks sheet, hover your mouse over any column header (in row 1) to see a pop-up note explaining what the column is for and what data is expected.

  - **Dropdown Lists:** Many cells have built-in dropdown lists to ensure you use standardised terminology (e.g., for samplingProtocol, sex, lifeStage). Please select a value from this list where available.

Warning: Be careful when pasting data from other sources. Pasting a value *into* a cell will overwrite and break the built-in dropdown lists and validation rules for that cell. The data will still

be validate when submitted, but the spreadsheet will no longer warn you of invalid data before submission.

- **Adding New Species to the Lists:** If you are reporting a host or vector species that is not in the dropdown list, you can overwrite the built-in validation by writing the name in a non-controlled cell, then copying and pasting. Alternatively, you can add it to the local reference list.

    1. Go to the 3.Vector species or 4.Host species worksheet.

    2. Insert a new row *in the middle* of the existing list.

    3. Type the new species name in the correct column.

    4. The new species will now be available in the dropdown on the data input sheet. *Note: When the validation server checks this new species, it will be flagged as a **Warning** (not an Error) because it is not in the system's main reference list. This is normal.*

2.3 How to Submit for Validation

1. Complete all your data entry in the **1.DATA INPUT-Ticks** worksheet.

2. Save the Excel file.

3. Create a new email and attach the saved file.

4. Send the email to the system's validation address: **vector.validation@efsa.epimundi.com**

2.4 Interpreting the Validation Response (Errors Found)

If the system finds any blocking errors in your file, it will send you a "Validation Failed" email. This email contains a summary of the issues and attaches two key files:

1. **The Annotated Spreadsheet (.xlsx):** This is a copy of your spreadsheet with automated feedback added. When you open it, you will see:

    o **Errors (Red Highlight):** These are critical issues that **must be fixed**. The data is invalid (e.g., a wrong date format, a missing required field).

    o **Warnings (Yellow Highlight):** These are potential issues that you **should check**. The data is valid but unexpected (e.g., a coordinate point is just outside the specified country, or you have used a species name not in the main database).

    o **Cell Comments:** Hover your mouse over any highlighted cell to read a specific comment explaining *exactly* what is wrong with that data.

2. **The Validation Map (.png):** This is a map that visually plots the coordinates of your records. It is used to help you spot spatial data errors (e.g., coordinates in the wrong country). Maps are only provided when there are spatial errors or warnings.

- **Understanding the New IDs:** In the annotated spreadsheet, you will see two new columns added at the end of your data:

- o **datasetName:** This is a unique ID for this specific submission attempt. It helps you track which version of the file you are working on.

- o **recordNumber:** This is a unique ID for each *row* of your data. This recordNumber is also plotted on the Validation Map, allowing you to easily find a problematic point from the map in your spreadsheet.

2.5 Correcting and Resubmitting

1. Open the annotated spreadsheet you received from the system.

2. Use the red highlights and cell comments to find and fix all **Errors**.

3. Review all **Warnings**. If the warning is for data that you have verified is correct (e.g., a new species or an unusual location), you may leave it.

4. Note – you don't have to worry about removing the highlight colours or notes. These are ignored when re-submitting the data.

5. Save the corrected file.

6. Send this new, corrected version back to the same validation email address.

You can repeat this cycle as many times as you need to fix all errors.

2.6 The "Success" Email: Approving for Final Submission

Once you submit a file that has **zero Errors**, the system will send you a "Validation Successful" email.

This email will provide:

- A summary of any remaining **Warnings** for you to review one last time.

- A unique, secure **"Submit to Review" link**.

This is the final step. When you are confident that your data is correct and that the remaining warnings are acceptable, click this link. This action gives your final approval and forwards your validated spreadsheet to the Reference Entomologist for manual review.

2.7 Receiving the Final File and Submitting to GBIF

After you click the "Submit to Review" link, your data is sent to the Reference Entomologist for their final scientific validation.

The entomologist will review your data (focusing on any warnings) and add their approval to each row in a new worksheet.

Once they have completed their review, they will **email this final, fully validated workbook back to you**. This file is now complete and ready for publication.

Your final responsibility as the data owner is to take this file and **submit it directly to the GBIF repository** according to the standard GBIF submission procedures. This completes the data validation and submission workflow.

2.8 How to Get a Customised Template

Users may submit data for a single country or multiple countries and require different optional columns. To make data entry simpler, you can generate a custom template containing only the reference data and columns you require for your purposes.

1. In the footer of any email you receive from the validation system (even a "fail" email), click the link "Request a Customised Spreadsheet Template."

2. This will open a secure web form in your browser.

3. On this form, you will:

   o Select the specific countries you are reporting on.

   o Identify the (single) vector group for the spreadsheet

   o Select any optional data columns you wish to add.

4. Click "Submit."

5. The system will generate a new template and email it to you. This template will have the NUTS3, Vector species, and Host species lists pre-filtered for only the countries and vector you selected, making data entry faster and cleaner.

Note: If you work with more than one vector group, download a different customised spreadsheet for each of those groups. This will allow the completed data to be sent to the appropriate reference entomologist for each group.

3 Guide for VectorNet Reference Entomologists

This section is for Reference Entomologists who perform the final scientific validation of a dataset.

3.1 Overview of Your Role

Your role is the final expert validation step before the data is submitted to GBIF.

The file you receive has already passed all automated technical checks, so it should contain **zero data Errors**. The data provider (submitter) has also already reviewed the file and confirmed that any remaining **Warnings** are acceptable to them.

Your job is to focus on the **scientific validity** of the data. You must pay special attention to the Warnings flagged by the system, such as new species, unexpected coordinates, or unusual species/protocol combinations, and use your expert judgment to approve or reject the records.

3.1.1 Workflow overview

- Submitter completes automated validation using the online tool
    - → Use online tool to submit to reference entomologist
    - → Reference entomologist checks relevant warnings and alerts
    - → Problems detected:
        - explain the problems, delete the GBIF submission page, and return to the submitter for correction
    - → No problems detected:
        - confirm validation (ORCID and '1' for each row) and return to the submitter for submission to GBIF

3.2 Receiving a Submission for Review

You will receive an email from the VectorNet Validation System, sent on behalf of the submitter, only *after* they have successfully validated their data and clicked the "Submit to Review" link.

This email will contain:

1. **The validated Excel workbook** as an attachment. This includes highlighting and annotation of any cells with warnings.

2. A list in the email body of any remaining **Warnings** that the automated system detected. This list helps you focus your review on the most critical data points.

3. A map may be attached to help identify any spatial errors or unusual findings

3.3 What to Check For

The data you receive may contain Errors, Warnings and Alerts

- **Errors**: You should *never* receive a data submission with errors. The submitter is required to correct these before submission.
- **Warnings**: These are unusual findings that could be errors. If you receive data with warnings, it means that the submitter should have already checked them and considers them to be correct.
  - These are often related to the NUTS unit in which the observation was made. As we assume the submitter has checked the warning, **you can generally ignore them**.
  - These highlights are retained in the spreadsheet so you can clearly see potential issues for follow-up if anything doesn't make sense.
  - Other warnings related to unusual (but not invalid) **species names**. You should **check these** and resolve them.
    - If there are valid species that are repeatedly causing warnings, you can add them to the standard reference list (see section 3.6 below).
- **Alerts**: These alerts are raised when the species reported is not recognised as endemic in the area in which it was reported.
  - You should check these carefully, as it may be due to either:
    - The movement of a species to a new area
    - An error in the identification of the species
    - An error in the data used to determine endemic and invasive species.
  - Note that the system may have problems matching species names when species complexes are used.

3.4 If problems are found: Providing feedback to the submitter

If errors are found, you should return the submission to the submitter by email, with an explanation of the problem. Include the note **Errors requiring correction** in the subject of the email.

- The submitter's email can be found in the first line of the email you receive from the system.
- You can explain the errors in any way that will be understood by the submitter:
  - In the text of the email
  - With manual highlights and comments in the data page (1.DATA INPUT) of the spreadsheet.
- You **should not** validate any of the data (add your ORCID) if there is invalid data
- You should **delete the GBIF submission page** (5.Data Output – GBIF). This ensures that the user cannot directly submit invalid data by mistake.

3.5 If no problems are found: The Validation Process

Open the attached Excel workbook. You will find that a **new worksheet** has been added to the file. This sheet is formatted in the required structure for the final GBIF submission.

Once you have checked all highlighted data and confirmed that it is all valid, you need to confirm validation in two columns. The column names are based on the GBIF standard, so they are not completely intuitive.

1. Enter your **ORCID** ID to the **identificationVerificationStatus** colum
2. Enter 1 (valid) in the **dynamicProperties** column
   a. If any specific rows are invalid, you can enter 0 in this column, but normally, all rows should be corrected by the submitter, so this should not happen.
3. Email the spreadsheet to the original data submitter (email address in the first line of the email that you received). They are then responsible for the final submission to GBIF

3.6 How to Add a New Species to the Reference List

As a Reference Entomologist, you have the authority to update the system's central reference lists for vector and host species.

You should do this when you receive a submission with a scientificName_invalid or associatedTaxa_invalid **Warning**. If you confirm that the species is scientifically valid but simply not in the system's database, you can add it permanently. This will prevent it from being flagged as a warning in future submissions.

3.6.1 Step 1: Access the Web Form

In the footer of any notification email you receive from the validation system, click the link labelled **"Update Species Reference List."**

This will open a secure web form in your browser.

3.6.2 Step 2: Fill Out the Form

The form is simple and requires three pieces of information:

1. **Species Name:** Type the full, scientifically correct name of the species you want to add (e.g., Aedes koreicus). You may also use synonyms or previous names

2. **Species Group:** Select whether this is a vectors or hosts species from the dropdown menu.

3. **GBIF Taxonomic Key:** This definitively associates the species with an item in the Darwin Core

3.6.2.1 A Note: How to Find the Correct GBIF Species Name

When you submit a new species, the system will query the GBIF database with the exact name you provide. To ensure the system finds the correct species and to prevent errors, we recommend you verify the spelling and find its official record on the GBIF website *before* submitting the form.

Here is the simplest way to find the correct record:

1. Go to the official GBIF species search page: **https://www.gbif.org/species/search**

2. Type the species name (e.g., Aedes koreicus) into the search box.

3. Click on the correct species from the list that appears in the main pane.

4. The page for that species will load. You can now do two things:

   o **Verify the Name:** The name on this page is the exact spelling the system will find.

- o **(Optional) Find the Key:** The GBIF taxonomic key is the number at the end of the URL in your browser's address bar. For example: https://www.gbif.org/species/**1234567**

By using the exact name from the GBIF site, you ensure that the system will automatically fetch the correct taxonomic record for the reference list.

3.6.3   Step 3: Submit

Click the "Submit" button.

The system will then perform several actions in the background:

1. It will take the species name you provided and query the **GBIF (Global Biodiversity Information Facility)** database.

2. It will retrieve the full, official taxonomic record for that species (including its GBIF key, phylum, class, family, etc.).

3. It will automatically add this complete, validated record as a new row in the system's central vectors.csv or hosts.csv reference file.

You will see a success message on the screen confirming the species has been added.

From that point forward, any future submission containing that species name will pass validation without a warning.

4 Guide for EFSA Managers

This section is for EFSA managers and administrative staff responsible for managing system access and monitoring activity.

4.1 Overview of Your Role

Your role is not technical but is critical for managing the system's security and monitoring its usage. Your two primary tasks are:

1. **User Management:** Maintaining the users.csv file, which controls exactly who is authorised to submit data to the system.

2. **Activity Monitoring:** Downloading and reviewing the logBook.csv file to see a high-level log of all submissions.

4.2 Access and Credentials

To perform these tasks, you must have secure access to the efsa.vector.ref S3 bucket in the AWS Management Console.

- **Request Credentials:** You must first request your access credentials from EpiMundi. You will be provided with a secure login URL, a username, and a password.

- **Security:** These credentials must be kept safe and confidential. They are tied to a specific IAM role that grants *only* the minimum permissions needed to upload the users.csv file and download the logBook.csv file. This role does not permit any other actions on the system.

4.3 Task 1: Managing the Authorised User List (users.csv)

The users.csv file is the master list that controls all system access. If a person's email is not in this file (or covered by a wildcard), their data submissions will be rejected.

4.3.1 Understanding the users.csv File Format

This is a simple Comma-Separated Values (CSV) file. You will edit it locally on your computer and then upload it to AWS to apply the changes.

The file has four columns. **You must not change the column names or order.** email,ms,role,vector

4.3.2 Column Definitions and Rules

- **email:**
  - **Single User:** user.name@example.com
    - Grants access to that specific email address.
  - **Wildcard Domain:** *@domain.com

- ▪ Grants access to *anyone* with an email from @domain.com.
- ▪ **Warning:** This should be used with extreme caution and only for small, fully-trusted institutions. It is always more secure to list individual email addresses.

- **ms:**
  - o This is the member state. Enter the two-letter code (in capitals) or set it to NA for users not associated with a single country

- **role:**
  - o Defines the user's permission level. This is
    - ▪ **admin** for EpiMundi,
    - ▪ **EFSA** for EFSA managers,
    - ▪ **ECDC** for ECDC access
    - ▪ **submitter** for data providers
    - ▪ **reference** for reference entomologists.

- **vector:**
  - o This column identifies the **Coordinating Reference Entomologists** who receive the final, validated submissions for a specific vector group.
  - o **Rule:** There must be **only two (2)** coordinating entomologists for each vector group.
  - o To assign a coordinator, enter one of the following values:
    - ▪ ticks,
    - ▪ mosquitoes,
    - ▪ midges,
    - ▪ sandflies.
  - o All other users—including data providers and other reference entomologists—must have this column left blank (or set to NA).

- **ohs**
  - o This column identifies if the user is a member of the One Health Surveillance project or not.
  - o Values:
    - ▪ 1 (is a One Health Surveillance project member)
    - ▪ 0 (is not a member)

4.4 How to Create and Save the File (Critical Formatting)

The system **requires** that the users.csv file is saved in **UTF-8 (without BOM)** format. A "BOM" (Byte Order Mark) is an invisible character that some programs (like Microsoft Excel) add to files, which will cause the system to fail.

4.4.1 Recommended Method (using Notepad++):

1. Open users.csv in a plain text editor like Notepad++ (free) or Visual Studio Code.

2. Make your changes (add or remove users).

3. Go to the **Encoding** menu.

4. Ensure that **"Encode in UTF-8"** is selected. (It must *not* be "UTF-8-BOM").

5. Save the file.

6. **How to Check:** When the file is open, the bottom-right status bar in Notepad++ should say **UTF-8**. If it says UTF-8-BOM, the file is incorrect.

4.4.2 Method 2 (using Microsoft Excel):

You can use Excel, but you must be careful when saving:

1. Make your changes in Excel.

2. Go to **File > Save As**.

3. In the "Save as type" dropdown, select **CSV UTF-8 (Comma delimited) (*.csv)**.

4. Save the file.

5. **Warning:** We still recommend checking the saved file in Notepad++ (as described above) to ensure no BOM was added.

4.5 How to Upload the users.csv File to AWS

1. Using the credentials provided by EpiMundi, log in to the AWS Management Console at the URL provided.

2. In the main console search bar, type **S3** and press Enter.

3. You will see a list of "Buckets". Click on the bucket named **efsa.vector.ref**.

4. You will see the list of all reference files. Find and click on the file users.csv.

5. You will see a "Versions" tab. Click it to see all previous versions of the file (your upload will be saved as a new version).

6. Go back to the "Objects" tab for the efsa.vector.ref bucket.

7. Click the **Upload** button.

8. Click **Add files** and select your newly saved users.csv file from your computer.

9. Click the **Upload** button at the bottom of the page.

10. The file will upload and *overwrite* the old file. The system will immediately begin using the new user list for all subsequent email submissions. The old file is preserved in the "Versions" tab if you need to revert.

4.6 Task 2: Checking the Activity Log

You can download a log of all submission activity at any time.

1. Log in to the AWS Management Console and navigate to the **efsa.vector.ref** S3 bucket, as described in steps 1-3 above.

2. Find the file named **logBook.csv**. This file contains a high-level record of all submissions, including the sender, timestamp, and whether the validation passed or failed.

3. Click the logBook.csv file name.

4. On the file details page, click the **Download** button.

5. You can open the downloaded CSV file in Microsoft Excel to review all system activity.

5 Appendix

5.1 Validation Errors

This table lists all critical errors that will block a file from being submitted. These issues must be fixed before the system will issue a "Validation Successful" email.

Field Tested	Category	Description of Check
1.DATA INPUT	structure	Missing required variables/columns
projectID	missing	Missing required value: please provide a unique project identifier
country	missing	Missing required value: please select a country name from the provided list
higherGeographyID	missing	Missing required value: please select a NUTS 3 code from the provided list
decimalLatitude	missing	Missing required value: please provide a value for the geographical latitude in the specified format.
decimalLongitude	missing	Missing required value: please provide a value for the geographical longitude in the specified format.
coordinatePrecision	missing	Missing required value: please select the level of precision for coordinates from the dropdown list.
CollectionEffortStartDate	missing	Missing required value: please provide a collection start date
CollectionEffortEndDate	missing	Missing required value: please provide a collection end date
samplingProtocol	missing	Missing required value: please select a sampling protocol from the dropdown list
sampleSizeUnit	missing	Missing required value: please select a category from the dropdown list
sampleSizeValue	missing	Missing required value: please provide a numeric value
scientificName	missing	Missing required value: please select a vector species from the dropdown list

decimalLatitude	datatype	The data type is not valid. Expected a numeric value.
decimalLongitude	datatype	The data type is not valid. Expected a numeric value.
coordinatePrecision	datatype	The data type is not valid. Expected a numeric value.
sampleSizeValue	datatype	The data type is not valid. Expected a numeric value.
individualCount	datatype	The data type is not valid. Expected an integer value (i.e., no decimals).
CollectionEffortStartDate	datatype	The data type is not valid. Expected a date value in excel date format (yyyy-mm-dd).
CollectionEffortEndDate	datatype	The data type is not valid. Expected a date value in excel date format (yyyy-mm-dd).
identifiedByID	datatype	The data type is not valid. The entered value doesn't fit the defined ORCID format, starting with "https://orcid.org". Also note that multiple DOIs should be separated by '
bibliographicCitation	datatype	The data type is not valid. The entered value doesn't fit the defined DOI format, starting with "https://doi.org/"
country	list	The submitted value is not valid. Please use only proposed values from the dropdown list.
higherGeographyID	list	The submitted value is not valid. Please use only proposed values.
coordinatePrecision	list	The submitted value is not valid. Please use only proposed values.
samplingProtocol	list	The submitted value is not valid. Please use only proposed values.
sampleSizeUnit	list	The submitted value is not valid. Please use only proposed values.
CollectionEffortStartDate	consistency	Invalid date value. The parsed date is in the future. Please check that the provided date is correct and in yyyy-mm-dd format.
CollectionEffortEndDate	consistency	Invalid date value. The parsed date is in the future. Please check that the provided date is correct and in yyyy-mm-dd format.

CollectionEffortStartDate	consistency	Please confirm that the date value is correct and in Excel date format (yyyy-mm-dd). The parsed date is before the year 1920.
CollectionEffortEndDate	consistency	Please confirm that the date value is correct and in Excel date format (yyyy-mm-dd). The parsed date is before the year 1920.
CollectionEffortStartDate	consistency	Invalid date value. The 'CollectionEffortEndDate' should be greater than 'CollectionEffortStartDate' (and vice-versa).
scientificName	consistency	The submitted vector species is not referenced in the GBIF vocabulary or is not a vector from the ticks, mosquitoes, sandflies or midges groups. Please confirm the provided vector species - if this is correct, please email your VectorNet contact point.
associatedTaxa	consistency	The submitted host is not referenced in the GBIF vocabulary. Please check the species. If the reference is correct, please email your VectorNet contact point.
country	consistency	The submitted coordinates (longitude and latitude) are not in the specified country. See provided map for more details. Please verify that the coordinates use the WGS84 reference system.
sex	list	The submitted value is not valid. Please use only proposed values from the dropdown list.
lifeStage	list	The submitted value is not valid. Please use only proposed values from the dropdown list.
occurrenceRemarks	list	The submitted value is not valid. Please use only proposed values from the dropdown list.
associatedTaxa	consistency	A host value was provided for a non-tick vector and a sampling protocol not using a bait.
associatedTaxa	missing	Missing required value: based on the indicated sampling protocol a host should have been specified.

5.2 Validation Warnings (Must be Checked)

This table lists all warnings. These will **not** block a submission, but they flag data that is unusual or unexpected. You and the Reference Entomologist must manually check these.

Field Tested	Category	Description of Check
verbatimSiteNames	datatype	The data type is not valid. The length of the text shouldn't be longer than three characters.
decimalLatitude	consistency	A numerical sequence of more than 5 elements was identified. Please confirm that this is not an input error due to value dragging.
decimalLongitude	consistency	A numerical sequence of more than 5 elements was identified. Please confirm that this is not an input error due to value dragging.
sampleSizeValue	consistency	A numerical sequence of more than 5 elements was identified. Please confirm that this is not an input error due to value dragging.
individualCount	consistency	A numerical sequence of more than 5 elements was identified. Please confirm that this is not an input error due to value dragging.
CollectionEffortStartDate	consistency	A numerical sequence of more than 5 elements was identified. Please confirm that this is not an input error due to value dragging.
CollectionEffortEndDate	consistency	A numerical sequence of more than 5 elements was identified. Please confirm that this is not an input error due to value dragging.
scientificName	list	The submitted vector is not a usual species identified through vector surveillance in the Vectornet Area (EU and neighbouring countries).
associatedTaxa	list	The submitted host is not a usual species identified through vector surveillance in the Vectornet Area (EU and neighbouring countries).
higherGeographyID	consistency	The submitted coordinates (longitude and latitude) are not in the specified NUTS. See provided map for more details.
1.DATA INPUT	columns	Some columns are not part of the expected columns (required or optional). These columns are ignored. If you would like these columns to be imported into the GBIF data, please contact biohaw@efsa.europa.eu to ask for a new optional column to be implemented.

5.3 External System Dependencies

The system's operation and maintenance are dependent on several external, third-party services.

- **Amazon Web Services (AWS)** The entire system is built and hosted on AWS and is fully dependent on its services for all operations. An outage in the eu-west-1 region for key services (S3, Lambda, SES, SQS) would result in partial or total system failure.

- **Global Biodiversity Information Facility (GBIF)** The system relies on the live GBIF API for the "Update Species Reference Data" workflow. If the GBIF API is unavailable, Reference Entomologists will be unable to add new species. This failure **will not** affect the primary data validation workflow for existing species.

- **Docker Hub (docker.io)** This service is a **build-time dependency**. The system's R container image is built using the rocker/r-ver:4.4 base image, which is pulled from Docker Hub. An outage at Docker Hub would prevent the build and deployment of new container versions.

- **Posit Public Package Manager (p3m.dev)** This is a **build-time dependency**. The Dockerfile installs all R packages from a specific p3m.dev repository snapshot. An outage at this repository would also prevent the build and deployment of new container versions.

5.4 Contact Information

5.4.1 Primary System Contact

For all general enquiries, technical support, or issues related to the data validation system, the primary point of contact is EFSA:

- **Email:** biohaw@efsa.europa.eu

5.4.2 Entomology and Data Enquiries

For specific scientific or entomological questions about data (e.g., how to classify a species, appropriate sampling protocols, or details on vector groups), please contact the relevant Vector Group Lead within the VectorNet network as listed on the GBIF VectorNet Portal.

Further information on the VectorNet project and its expert network can be found on the GBIF VectorNet Portal website https://www.vectornetdata.org/.