

## SHORT TERM SCIENTIFIC MISSION (STSM) SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

**Action number:** COST Action CA17110

**STSM title:** WG2, STSM4 Existence and quality of data about control programmes for cattle diseases

**STSM start and end date:** 01/09/2020 to 30/09/2020

**Grantee name:** Eglė Rapaliūtė

### **PURPOSE OF THE STSM:**

(max.200 words)

Following the results of the STSM2 and STSM3: a tool for data quality evaluation and Limesurvey questionnaire "SOUND-control WG2 questionnaire", the main purpose of STSM4 was to analyse the data collected by the questionnaire and provide an overview about the existence and quality of data about cattle industry and control programmes (CPs) that could be used to estimate disease freedom in the different partner countries of SOUND control for a range of cattle infectious diseases.

The results of data analysis will capture the heterogeneities of the existing data among Action member countries and will also answer several critically important questions of WG2 and the SOUND control: 1) Are there enough data that could be used to estimate disease freedom in all Action countries?; 2) Is the quality of data poor/fair/good among Action member countries (quality in terms of accessibility, completeness, timeliness and accuracy) (Annex 1)?; 3) How many and what kind of sources members need to use to obtain the required data?

Moreover, it was important to design sustainable data analysis workflow (i.e. R studio script) that the results of data analysis could be possibly updated on a regular basis, as more countries can join the Action and complete the questionnaire.

### **DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS**

(max.500 words)

#### 1<sup>st</sup> week (01/09 – 04/09):

The WG2 questionnaire was designed using "Limesurvey" software, provided by Veterinary Public Health Institute (VPHI), University of Bern. By the first day of STSM4, 20 of 32 SOUND control members, representing their countries completed the questionnaire. The response rate was evaluated and additional reminders were sent to 7 SOUND control members, who previously received an invitation to participate and 5 invitations to new contacts were sent to countries with no response to the first invitation. Therefore, by the end of STSM 4 the response rate was the same, but at least 3 countries are in progress of completing the questionnaire and WG2 members keep on working to get in contact with Action member countries with no response, the results will be updated regularly.

#### 2<sup>nd</sup> week (07/09 – 11/09):

The data of the questionnaire was extracted from "Limesurvey" software and explored by evaluating the types of data, completeness of answers, missing values, presence of possible duplicates or any possible data extraction errors and describing the definitions of dairy and beef cattle. The plan of data analysis was additionally discussed with the supervisor Dr Lena-Mari Tamminen and decided to analyse data in parts for: data existence, data quality (overall and by every criterion) and the summary of sources used to obtain the data. The questionnaire for data analysis was subseted in 3 parts: 1) Demographics, 2) Risk factors, 3) Disease control programmes (3.1. Bovine viral diarrhoea (BVD), 3.2. Infectious Bovine rhinotracheitis (IBR)

and 3.3. Johne's disease (JD)) and the names of the variables changed if needed. The definitions of dairy and beef cattle varied between countries and very little data was provided, it was decided to only analyse the data about the total cattle population (dairy and beef together).

3<sup>rd</sup>, 4<sup>th</sup> week (14/09 – 18/09, 21/09 – 25/09):

Data was analysed using R Studio software. The main packages used for data analysis were for data handling, manipulation, descriptive statistics and visualization: dplyr, tidyverse, ggplot2 etc.

The data was analysed according to the plan: 1) Demographics, 2) Risk factors and 3) Disease control programmes (BVD, IBR, JD): data existence, the quality of data (overall and by every criterion) and summary of sources used to obtain the data. For the visualization of the results, it was decided to mainly use heat maps (ggplot2). The interpretation and description of data analysis results were also carried out, by preparing an additional document "Existence and quality of data about cattle demographics, risk factors for disease introduction and control programmes (JD, IBR, BVD) in SOUND – control member countries".

5<sup>th</sup> week (28/09 – 30/09):

The completion of data analysis and the first version of document describing the analysis results was prepared. Preliminary information about the completeness of the questionnaire and the results of data analysis were discussed on the 30<sup>th</sup> of September, during WG2 telco meeting. Also, during the last week of STSM the publication for Frontiers in Veterinary science – Veterinary Epidemiology and Economics, special issue: Global Control and Eradication Programmes For Cattle Diseases, was prepared.

### **DESCRIPTION OF THE MAIN RESULTS OBTAINED**

The main results of data analysis will be described in a document prepared during the STSM4: "Existence and quality of data about cattle demographics, risk factors for disease introduction and control programmes (JD, IBR, BVD) in SOUND – control member countries" and also used in WG2 deliverable. Nevertheless the document with the results will be firstly approved by WG2 and SOUND control members before publicizing. Also, R studio script is prepared and can be used after new countries complete the questionnaire. The main parts of the results can be listed as following:

1. Data existence in SOUND – control member countries:

1.1. Existence of data about the demographics of cattle population

1.2. Existence of data about risk factors of disease introduction to the herd

1.3. Existence of data about the disease control programmes:

1.3.1. Johne's disease (JD), 1.3.2. Infectious Bovine Rhinotracheitis (IBR), 1.3.3. Bovine Viral Diarrhoea (BVD).

2. Data quality evaluation in SOUND – control member countries:

2.1. Quality evaluation of demographics data

2.2. Quality evaluation of risk factors data

2.3. Quality evaluation of disease control programmes data:

2.3.1. Johne's disease (JD), 2.3.2. Infectious Bovine Rhinotracheitis (IBR), 2.3.3. Bovine Viral Diarrhoea (BVD).

3. Sources used to obtain data:

3.1. Sources used to obtain the demographics data

3.2. Sources used to obtain the risk factors data

3.3. Sources used to obtain the disease control programmes data.

### **FUTURE COLLABORATIONS (if applicable)**

WG2 STSM4 grantee Ms Rapaliute together with the supervisor of STSM3 Dr Céline Faverjon, supervisor of STSM4 Dr Lena-Mari Tamminen in collaboration with WG2 and SOUND control members will prepare a publication describing the outcomes of the WG2 questionnaire analysis. The publication will be submitted to Frontiers in Veterinary science – Veterinary Epidemiology and Economics, special issue: Global Control and Eradication Programmes For Cattle Diseases.

Annex 1

Table 1. Data quality evaluation tool (designed during WG2 STSM3)

| Quality criteria<br>Evaluation | Accessibility (1)   | Completeness (2)  | Timeliness (3)                             | Accuracy (4)  |
|--------------------------------|---|---|--|---|
| <b>POOR</b><br>Score - 1       | The variable is not routinely collected <b>AND</b> you only have access to this information via indirect sources (e.g., research studies)   | The variable is not mandatory to fill in the data base <b>AND</b> completeness of data is unknown <b>OR</b> lower than 80%                    | It is unknown when data is updated         | Variable is entered manually to the dataset <b>AND</b> No data validation is performed (e.g., the data are not used for any purpose).   |
| <b>FAIR</b><br>Score - 2       | The variable is not readily available but can be obtained by combining multiple sources <b>AND/OR</b> data is available, but access is associated with fee/approval of data-owner | The variable is not mandatory to fill in the data base <b>AND</b> completeness of data set is >80 %   | The data is updated once or twice per year | Variable are entered manually <b>AND</b> data validation procedure is sometimes implemented (e.g., variable is used on a regular basis for creating reports, or combined with other data sources)   |
| <b>GOOD</b><br>Score - 3       | The variable is obtained from one data source <b>AND</b> can be extracted when needed   | The variable is mandatory to fill in <b>OR</b> The variable is not mandatory to report, <b>AND</b> completeness of data set is close to 100 % | The data is updated real time              | The variable is collected and entered by automatic system/robot <b>OR</b> The variable is entered manually <b>AND</b> data validation procedure is always implemented (e.g., variable is used on a regular basis for creating reports, or combined with other data sources) |

Table 2. Overall data quality evaluation (designed during WG2 STSM3)

|                              |  |
|------------------------------|--|
| Score 1-4<br>Poor quality    | Very little information available of this variable, it is not mandatory to fill in data set and there are a lot of missing or unknown values. The data set of this variable is rarely updated, and no data validation is performed.  |
| Score 5 – 8<br>Fair quality  | Variable is available but not readily accessible. The variable is not mandatory to fill but it is possible to check the completeness of data set and it is more than 80%. The variable is regularly used, and some data validation is implemented.   |
| Score 9 – 12<br>Good quality | A lot of information of this variable is available and it is easy to extract it when needed. The variable is mandatory to fill in and it is possible to check the completeness of data automatically, which is more than 80%. A data is updated often, it is collected by robot or manually, but validation procedure is always implemented. |