The Zoonoses Data Collection in Italy: An Expert System for Data Quality Management and Improvement

Iannetti S[1]*, Cioci D[2], Falcone MG[3] and Colangeli P[2]

[1]*Veterinary Medicine, Istituto Zooprofilattico Sperimentale of Abruzzo and Molise G. Caporale, Campo Boario Teramo, Italy*
[2]*Computer Science, Istituto Zooprofilattico Sperimentale of Abruzzo and Molise G. Caporale, Campo Boario Teramo, Italy*
[3]*Veterinary Medicine, Ministry of Health, Office III DGSAF, Rome, Italy*

Abstract

The need to improve the quality of data for a better analysis and understanding of zoonoses' trend at country level has been increased year by year both by the EFSA and by reporting countries. In the framework of an EFSA's Grant project, aimed to complement the zoonoses historical database, an expert system based on logical rules of truth tables was put in place within the Italian information system for zoonoses data collection (SINZoo). During data entry, the truth tables check that, for each zoonoses, the combination of the area of interest, each possible sampling context, stage and sampling unit has been entered correctly, thus avoiding inconsistent data. Each combination available in the truth tables indicates the context, the stage, the sampling unit allowed for each zoonosis in a specific area and for a category of species. The goal of the project was achieved for most of the information to be retrieved: the 89% and the 83% of sampling contexts and stages respectively and the 100% of the other information were retrieved. To date, the truth tables developed for specific zoonoses have become integral part of SINZoo, allowing to avoid mistakes during data reporting. Data quality is the pillar for any analysis and to perform risk analysis: the logical rules of truth tables can be implemented in other information systems involved in data collection in the field of animal health and food safety, increasing both the consistence and the coherence of the data reported.

Keywords: Business rules; Data collection; Data quality; Data dictionary; Expert system; Information system; Zoonoses

Introduction

The European Union (EU) system for monitoring and collecting information on zoonoses is based on Directive (EC) n. 2003/99, which obliges the EU Member States to collect data on the occurrence of zoonoses, zoonotic agents, antimicrobial resistance, animal populations and food-borne outbreaks [1]. Moreover, under Regulation (EC) n. 178/2002, the European Food Safety Authority (EFSA) is assigned the tasks of examining these data and publishing annual European Union Summary Reports (EUSR) in cooperation with the European Centre for Disease Prevention and Control (ECDC), which provides and analyses the data on zoonotic infections in humans [2]. In order to collect data by the different Member States, which already have different systems to collect zoonoses data, EFSA developed a Zoonoses Data Model and a common dictionary (contained in pick lists) to be used by each reporting country, based on the Standard Sample Description, for all the items to be reported, with the aim of collecting data in a unique and uniform way and with the same semantic [3].

In Italy, the Italian Ministry of Health (MoH) is the national authority in charge of collecting data on zoonoses and zoonotic agents covered by Directive (EC) n. 2003/99 [4]. Since 2008, zoonoses data collection takes place at national level through the "national information system for zoonoses data collection", named SINZoo, developed in compliance with the Directive (EC) n. 2003/99 by the Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale" (IZSAM) on request of the MoH. Zoonoses data are registered into SINZoo by the regional veterinary services through different alternatives: on line forms, upload of XML files and Web Services [5]. A common dictionary was developed in SINZoo, integrated with the EFSA's one and with other categorizations specifically developed at national level. The database in SINZoo was established considering both the information debt toward the EFSA and the national needs and is made up of the following components:

- Zoonosis

- Tested species for animal prevalence data, species of origin of the food for food prevalence data, species of destination of feed for feed prevalence data

- Sampling Context

- Sampling Stage/Sample Type

- Sampling Unit

Each component includes a list of possible values which found the correspondence (mapping) of the values in the EFSA's catalogue [6].

SINZoo is strictly integrated with other national information systems already existent in animal health and food safety and is available on the website of the National Veterinary Information systems, thus avoiding double data entry and assuring data uniqueness and the coherence between dictionaries and definitions [7-9]. In fact, the national information debt towards the EFSA consists in aggregated data (i.e., not sample based) which come from the different national information systems:

- Data on animal population, available in the National veterinary database for livestock and holdings (BDN),

- Data on zoonoses and zoonotic agents in animal, food and feed,

stage should be "at slaughterhouse", while in wildlife the sampling stage should be "at game handling establishments".

Data on *Toxoplasma*, *Yersinia*, *Staphylococcus*, *Campylobacter* and *Coxiella* in animals are mainly collected during clinical investigations, or surveys, since for the years covered by the project there was not a national monitoring system in place for these zoonoses in Italy. Therefore for such zoonoses the only possible sampling context in the EFSA's historical database could be "survey" or "clinical investigation". Clinical investigations are usually carried out for testing of symptomatic animals suspected to be affected by any zoonotic agent, therefore the sampling strategy was updated to "suspect sampling" when missing. The sampling context was also revised in case of mistakes: the main corrections regarded for instance official sampling activities reported wrongly with sampling context "monitoring" or "surveillance" for which the sampling context was amended and the sampling strategy was corrected to "objective sampling" or to "suspect sampling" in case of surveys or clinical investigations respectively.

Examples of logical rules developed for missing, unspecified, wrong sampling stage

The sampling stage was updated to "at farm" in case of farmed animals or of zoo animals, (since this stage in SINZoo was intended to be used for all type of premises keeping animals). As regards pets, in Italy most of them are investigated at private level; to date, EFSA does not include this stage in the pick lists, so it was left to "unspecified" in the EFSA's historical database. For wild animals, the only possible sampling stages could be "natural habitat", "conservation facilities" or "game handling establishments", depending on the zoonosis, the tested species and other rules described previously.

Examples of logical rules for generic species

EFSA implemented the matrix pick list within its catalogue, asking Member States to specify if a given species was farmed or wild, in order to obtain more and more detailed data. Therefore, a logical rule was

implemented for wild boars and water buffaloes to decide whether the animals sampled were farmed or not. So, in case of water buffalo, this species was updated to "water buffalo-farmed" and the stage was updated to "at farm" when unspecified or missing, since this species in Italy is only farmed (Table 1). For *Echinococcus* and *Trichinella* in water buffaloes and in wild boars-farmed, the only possible sampling stage could be "at slaughterhouse", since according to the legislation, these species shall be systematically sampled at the slaughterhouse [16], while for wildlife-wild boars, the only possible sampling stage could be "natural habitat", where the animals can be hunted (Table 2). The correction of the reported species regarded "pigs-mixed herds-sows", which were corrected to "pigs - breeding animals - unspecified – sows" and "pigs-mixed herds-fattening pigs" were corrected to "pigs - fattening pigs – unspecified", since there are not mixed herds of farmed pigs in Italy.

Results

The logical rules described previously were used to create a truth table for each zoonosis to be reported to the EFSA. Each truth table was implemented taking into account the relationship between the zoonosis, the area of interest (animal, food, and feed) and each possible sampling context, stage and sampling unit. From Decoding menu of SINZoo, it is possible to download truth tables for each of the zoonoses to be reported to the EFSA according to the Dir. 2003/99/CE (Figure 2). The entire EFSA's historical database was processed by the truth tables in order to extract the rows containing combinations fields which were not admitted. Each time one or more row was updated, the system verified its correctness by checking the truth table for the zoonosis in question.

The database of SINZoo is logically structured within the entity-relationship (E/R) schemas (Figure 1). The truth table is linked, through foreign keys, to:

• zoonosis,

COD_EFSA	DESCR_EFSA	stage	stage update to:	species updated to:	species considered as:
A003581A	Wildlife - wild boars	unspecified or null	natural habitat		wild animals
A020861A	Wild boars - farmed	unspecified or null	at farm		Farmed/zoo animals
A010041A	Wild boars	at farm		wild boars - farmed	Farmed/zoo animals
A010041A	Wild boars	unspecified or null	natural habitat	Wildlife - wild boars	wild animals
A003581A	Wildlife - wild boars	at farm	natural habitat		wild animals
A020861A	Wild boars - farmed	natural habitat	at farm		Farmed/zoo animals
A006821A	Water buffalos	unspecified or null	at farm	Water buffalos - farmed	Farmed/zoo animals
A006821A	Water buffalos	at farm		Water buffalos - farmed	Farmed/zoo animals
A031641A	Water buffalos - farmed	unspecified or null	at farm		Farmed/zoo animals
A031641A	Water buffalos - farmed	natural habitat	at farm		Farmed

Table 1: Logical rules followed to update generic species.

COD_EFSA	DESCR_EFSA	stage	stage update to:	species updated to:	species considered as:
A003581A	Wildlife - wild boars	unspecified or null	natural habitat		wild animals
A003581A	Wildlife - wild boars	at farm	natural habitat		wild animals
A010041A	Wild boars	at farm	at slaughterhouse	wild boars - farmed	Farmed/zoo animals
A010041A	Wild boars	unspecified or null	natural habitat	Wildlife - wild boars	wild animals
A020861A	Wild boars - farmed	unspecified or null	at slaughterhouse		Farmed/zoo animals
A020861A	Wild boars - farmed	natural habitat	at slaughterhouse		Farmed/zoo animals
A006821A	Water buffalos	unspecified or null	at slaughterhouse	Water buffalos - farmed	Farmed/zoo animals
A006821A	Water buffalos	at farm	at slaughterhouse	Water buffalos - farmed	Farmed/zoo animals
A031641A	Water buffalos - farmed	unspecified or null	at slaughterhouse		Farmed/zoo animals
A031641A	Water buffalos - farmed	natural habitat	at slaughterhouse		Farmed/zoo animals

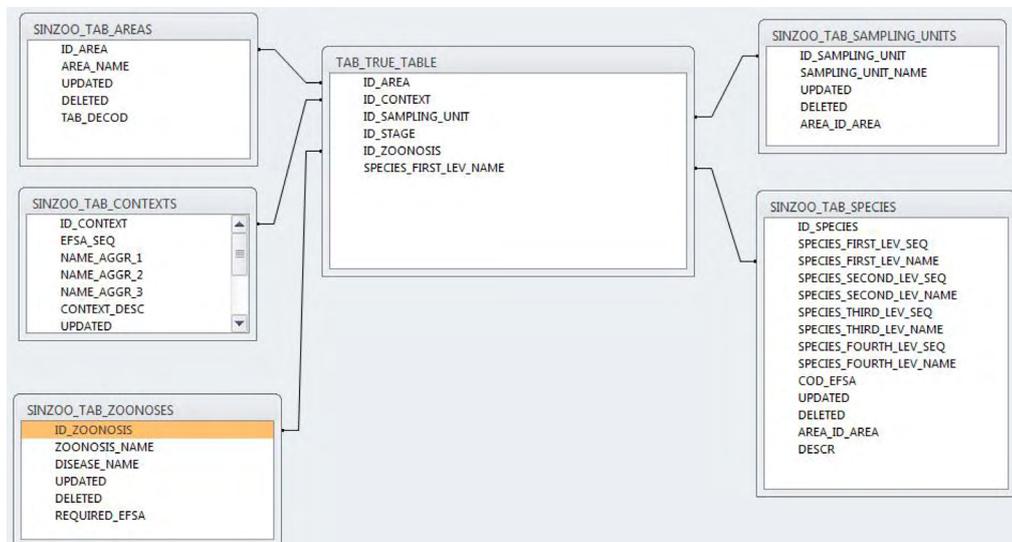**Table 2:** Logical rules followed to update generic species in case of *Echinococcus* and *Trichinella*.

Figure 1: Entity-relationship (E/R) scheme.

Figure 2: Truth table trichinosis.

- area,

- context,

- stage,

- sampling unit,

- Species (first level name): this field stands for the animal species if the reporting area is "Animal", the species of origin of the food if the area is "Food" and the species of destination of the feed if the area is "Feed".

Each combination available in the truth table indicates the context, the stage, the sampling unit allowed for the reported zoonosis in a specific area and for a category of species. Thanks to the logical rules and the truth tables, the missing/unspecified information was retrieved and the rows containing wrong combinations of zoonosis/ context/ stage/unit were corrected and updated. Overall, the goal of the Grant project was achieved for most of the information to be retrieved: the 89% and the 83% of sampling contexts and stages respectively and the

100% of the other information were retrieved (Table 3). The entire process was realised following logical rules both of the truth table and of the national legislation in place for the zoonoses covered by the project, thus ensuring the coherence of all the data retrieved. Also, the whole work was made easier by the implementation of the web interface shown in Figure 3: an alert "message" notified if the record selected had different or equal values for the fields composing the "context", i.e., sampler/ sampling context/ sampling strategy and what type of values. The user might change the selected values and by clicking on "Confirm select", saved the modification. Contextually, the system processed new combination and re-generates the comment field with success/error/ warning (Figure 3).

Examples of truth table developed for some specific zoonosis and zoonotic agent

Toxoplasma, Yersinia, Staphylococcus, Campylobacter, Coxiella: National monitoring programme for *Toxoplasma, Yersinia, Staphylococcus, Campylobacter* and *Coxiella* in animals are not in place

Reporting Year	Sampling Context	Sampler	Sampling Strategy	Sampling Stage	Sample Type	Sampling Unit
2008	100%	100%	100%	80%	100%	-
2009	99%	100%	100%	86%	100%	-
2010	100%	100%	100%	80%	100%	100%
2011	87%	-	-	87%	100%	-
Total	**89%**	**100%**	**100%**	**83%**	**100%**	**100%**

Table 3: Percentage of retrieved information for each SCs.

Figure 3: Web inerface.

for susceptible species, so most of the sampling activities are carried out during surveys or clinical investigations in animals suspected to be diseased, therefore for such zoonoses the truth tables admit only two sampling context (and only the susceptible species), "survey" or "clinical investigation". Moreover, in case of Campylobacter in poultry and birds, Campylobacteriosis does not cause clinical signs, therefore the truth table does not admit "clinical investigations" in these species. The sampling strategy admitted for clinical investigation is "suspect sampling", since the animals are controlled in case of suspectd disease. Also sample type admitted are coherent with the zoonosis: i.e., for Toxoplasma, typically blood (for intermediate hosts) is tested by serology, other samples could include abortion material (e.g., sheep) or faeces (e.g., cats); for Yersinia and Campylobacter, also faeces and environmental samples are admitted in the truth tables.

Echinococcus: Echinococcosis is one of the diseases listed in the Annex I part A, therefore it shall be monitored by the official competent authorities. In Italy, surveillance is performed during official meat inspection as part of Regulation (EC) No 854/2004 of the European Parliament and of the Council of 29th April 2004 laying down specific rules for the organisation of official controls on products of animal origin intended for human consumption [17]. Some regional programmes are in place on the national territory and official sampling activities take place at slaughterhouse as part of the post-mortem examination for all farmed animals. Therefore for echinococcosis the truth table admits the context "monitoring", sampler "official sampling", sampling strategy "census" and sampling stage "at slaughterhouse", according to Regulation (EC) No. 854/2004. The sample type admitted is "animal sample-organ/tissue". The sample unit is always "animal" since the unit tested in the framework of the monitoring programme is always the animal.

Brucella: The truth table for brucellosis does not admit official sampling activities carried out on cattle, sheep and goats and farmed water buffaloes, since these data are provided by the Italian information system on Community co-financed eradication programmes (SIR). Clinical investigations are admitted for other susceptible species, with sampling strategy "suspect sampling": for instance, pets may be subjected to clinical investigations carried out in case of suspect of disease, especially for dogs living on a farm. The sampling context "survey" is admitted for animal species susceptible to Brucella, with sampler "objective sampling". For sample type, typically blood is tested by serology; other animal samples could include abortion material. Sampling stage "natural habitat" is admitted for wild animals and "at farm" for farmed or zoo animals.

Trichinella: According to Regulation (EC) n. 2075/2005, the competent authority should implement a monitoring programme covering domestic swine, horses and other susceptible animal species coming from holdings or categories of holdings recognised as free from *Trichinella* or from regions where the risk of *Trichinella* in domestic swine is recognised as negligible, in order to verify that the animals are effectively free from *Trichinella*. To that end, meat samples shall be collected and examined for presence of Trichinella parasites. The truth table developed for trichinosis in animals admits only "animal" as sampling unit, "monitoring" as sampling context, "official sampling" as sampler and "census sampling" as sampling strategy. Since trichinosis is asymptomatic in most of the animal species, the context "clinical investigations" is not admitted.

Discussion

Data reported by the Member States and other reporting countries annually are stored in the EFSA's Zoonoses databases and used for the production of the annual European Union Summary Report (EUSR), other scientific or technical reports [18,19], and for supporting risk analysis carried out by EFSA's scientific panels [20,21]. Since changes in the historical datasets of Member States occurred over the years and the standardisation of the data reporting improved during the past years, needs to update and complement the national historical datasets in the EFSA's Zoonoses databases were required. This was to guarantee the correctness of the data and the subsequent accuracy of the analyses based on the national data. The work done during the Grant project permitted to trace, wherever possible, the missing right information and to improve, overall, the quality of data reported to the EFSA from 2008 to 2011. The need to retrieve historical data led to improve the rules of the national data collection foreseen by the truth tables and to define new logics and algorithms that may be used and improved for each reporting.

In conclusion, the work done during the project permitted to trace, wherever possible, the missing right information and to improve, overall, the quality of data reported to the EFSA from 2008 to 2011. The Italian team developed an expert, flexible system, using business rules (i.e., truth tables) able to support the annual collection both at aggregate level and sample based level. From the Grant project until the last reporting season (2015 zoonoses data collection) the implementation of the truth tables in SINZoo allowed checking the correctness of the data reported, thus avoiding rough mistakes during the system's feeding. To date, the truth tables contain all the valid combination of context, stage, sampling unit allowed for a given zoonosis in a given area (animal, food, and feed) and for a category of species. During zoonoses data entry, the truth table checks that, for each zoonosis, the combination of the area of interest, each possible sampling context, stage and sampling unit has been entered correctly, thus avoiding inconsistent data. Moreover, the truth tables are under revision annually by the national Panel of expert: the revision takes into account any possible change in the national or EU legislation and also the addition of new combinations for which the experts decide the cogency.

In the absence of any structured information system, data collection would be prone to delays, errors and omissions. In fact, the same information could be requested several times and in different formats, causing confusion for the lack of a unique data collection.

This Grant project highlighted the importance of data quality during the collection and feeding of any information system. Data quality is the pillar for any epidemiological analysis and to perform risk analysis: logical rules of truth tables can be implemented in other information systems involved in data collection in the field of animal health and food safety, increasing both the consistence and the coherence of the data reported.

Ensure the quality of the data collected is the basis for any subsequent processing. For this reason, the efforts go towards the integration of the various information systems, avoiding duplication of the same data in different systems. The data is entered only in one spot, usually by those who created it and is made available to those who need it (interoperability). Moreover, the existing national information systems shifted their ability over time from operational applications (collecting data in course of normal business operation) to decision support systems having primarily the purpose to collect appropriate data of high quality and presenting the results of data analysis to stakeholders and decision makers.

References

1. European Parliament, Council of the European Union (2003) Directive 2003/99/EC of the European Parliament and of the Council of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC, pp: 31-40.

2. European Parliament, Council of the European Union (2002) Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety, pp: 1-24.

3. EFSA (European Food Safety Authority) (2010) Standard sample description for food and feed. EFSA Journal 8: 1457.

4. Legislative Decree No. 191 of 4th April (2006) Implementation of Directive 2003/99/EC on the monitoring of zoonoses and zoonotic agents. Italian Official Gazette no. 119 of 24 May 2006.

5. Colangeli P, Iannetti S, Ruocco L, Forlizzi L, Cioci D, et al. (2012) The Italian Information System on Zoonoses Data Collection. Zoonoses Public Health 60: 182-188.

6. European Food Safety Authority (EFSA) (2014) Manual for reporting on zoonoses and zoonotic agents, within the framework of Directive 2003/99/EC and on some other pathogenic microbiological agents for information deriving from the year 2015. 573: 107.

7. Colangeli P, Iannetti S, Cerella A, Ippoliti C, Di Lorenzo A, et al. (2011) The National information system for the notification of animal diseases in Italy. Vet Ital 47: 303-312.

8. Iannetti S, Malizia E, Sorgente A, Ricci A, Cibin V, et al. (2015) The Integrated Management System for Salmonella Control Plan in Italy. J Health Med Informat 6: 199.

9. Napoli C, Iannetti S, Rizzo C, Bella A, DiSabatino D, et al. (2014) Vector borne infections in Italy: results of the integrated surveillance system for West Nile Disease in 2013. BioMed Res International. 2015: Article ID 643439.

10. Commission Decision No. 2007/851/EC of 10 December 2007 amending Decisions 2006/687/EC, 2006/875/EC and 2006/876/EC as regards the reallocation of the Community's financial contribution to certain Member States for their programmes for the eradication and monitoring of animal diseases and for checks aimed at the prevention of zoonoses for 2007 (notified under document number C (2007) 5985).

11. Commission implementing Decision No. 2011/807/UE of 30 November 2012 approving annual and multiannual programmes and the financial contribution from the Union for the eradication, control and monitoring of certain animal diseases and zoonoses presented by the Member States for 2013 and following years.

12. Commission Regulation (EC) No. 1003/2005 implementing Regulation (EC) No. 2160/2003 of the European Parliament and of the Council as regards a Union target for the reduction of the prevalence of certain Salmonella serotypes in breeding flocks of Gallus gallus and amending Regulation (EC) No. 2160/2003.

13. Commission Regulation (EU) No. 1190/2012 of 12 December 2012 concerning a Union target for the reduction of Salmonella Enteritidis and Salmonella Typhimurium in flocks of turkeys, as provided for in Regulation (EC) No 2160/2003 of the European Parliament and of the Council Text with EEA relevance.

14. Commission Regulation (EC) No. 646/2007 of 12 June 2007 implementing Regulation (EC) No 2160/2003 of the European Parliament and of the Council as regards a Community target for the reduction of the prevalence of Salmonella Enteritidis and Salmonella Typhimurium in broilers and repealing Regulation (EC) No 1091/2005.

15. Commission Decision No. 2007/848/EC of 11 December 2007 approving certain national programmes for the control of Salmonella in flocks of laying hens of Gallus gallus.

16. Commission Regulation (EC) No. 2075/2005 of 5 December 2005 laying down specific rules on official controls for Trichinella in meat.

17. Regulation (EC) of the European Parliament and of the Council No. 854/2004 (EC, 2004) of 29th April 2004 laying down specific rules for the organisation of official controls on products of animal origin intended for human consumption.

18. EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control) (2016) The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2015. EFSA Journal. 14: 04634.

19. European Food Safety Authority (EFSA), Boelaert F, Hugas M, Ortiz Pelaez A, Rizzi V, Stella P et al. (2016) The European Union summary report on data of the surveillance of ruminants for the presence of transmissible spongiform encep halopathies (TSEs) in 2015. EFSA Journal. 14: 4643.

20. Opinion of the Scientific Panel on Biological Hazards on "Risk assessment of a revised inspection of slaughter animals in areas with low prevalence of Trichinella". The EFSA Journal 200: 1-411.

21. Scientific Opinion of the Panel on Biological Hazards on a request from EFSA on Surveillance and monitoring of Toxoplasma in humans, foods and animals (2007) The EFSA Journal 583: 1-64.