

APPROVED: 30 May 2018

doi:10.2903/sp.efsa.2018.EN-1428

SIGMA

A comprehensive animal disease data collection approach

Harmonised data model for domestic and wild animal populations

European Food Safety Authority

Abstract

The European Commission is asking EFSA for scientific and technical support in the epidemiological analysis of animal disease outbreaks (i.e. African swine fever, lumpy skin disease and avian influenza) and to assess surveillance data (i.e. *Echinococcus multilocularis* and avian influenza). At the moment, since EFSA data models are not completely harmonised; EFSA's data warehouse is not connected with the EC's Animal Disease Notification System (ADNS); and data submission and validation is not automated, the animal health data collections are labour intensive for both EFSA and the data providers of the Member States. EFSA's aim is to work together with Member States on the technical aspects of 'ad hoc animal health data collections', in order to (i) reduce the manual input of the data to be submitted by the Member States to EFSA; (ii) avoid double reporting to EFSA and, possibly, to other systems; (iii) Provide the Member States with instruments to produce automatically draft national reports on animal health and surveillance in a protected environment (secure connections, log in credentials) to ensure data protection; (iv) increase the quality and the comparability of the data received from the Member States; (v) shorten the time to retrieve up-to-date data, relevant for risk assessment purposes, and to release sound scientific advice. This report aims at providing the reader with the main elements to understand the SIGMA project and at briefly describing the contents of the Animal Health Network meeting and at capturing the feedbacks having a direct impact on the strategy.

© European Food Safety Authority, 2018

Key words: SIGMA, data collection, harmonisation, standardisation, support

Question number: EFSA-Q-2018-00452

Correspondence: any enquiries related to this output should be addressed to ALPHA@efsa.europa.eu

Disclaimer: The views or positions expressed in this publication do not necessarily represent in legal terms the official position of the European Food Safety Authority (EFSA). EFSA assumes no responsibility or liability for any errors or inaccuracies that may appear.

Acknowledgements: EFSA wishes to thank the members of the Working Group on SIGMA: Helen Roberts, Marta Bedriova, Søren Nielsen Saxmose, Hans-Herman Thulke and Paolo Calistri for contributing to the development of the SIGMA ADM; the members of the Animal Health Network: Robert Wolf, Kirstine Ceulemans, Lilyana Polihronova, Drazen Knezevic, Georgios Krasias, Richard Wallo, Anette Boklund, Ave Ly Toomvap, Pascal Hendrikx, Aik Plevraki, Melinda Kocsis, Ronan O’neill, Fabrizio De Massis, Edvins Olsevkis, Olaf Stenvers, Dean Bas, Przemyslaw Cwynar, Yolanda Vaz, Anna Ondrejкова, Luis Romero, Cecilia Hulten, Helen Roberts for participating to the annual meeting; and EFSA staff members: Denise Candiani, Sotiria-Eleni Antoniou, Frank Verdonck, Alessandro Broglia, Mario Monguidi and Gabriele Zancanaro, for the support provided to this scientific output.

Suggested citation: EFSA (European Food Safety Authority), 2018. SIGMA - A comprehensive animal disease data collection approach. EFSA Supporting publication 2018:EN-1428. 7 pp.

ISSN: 2397-8325

© European Food Safety Authority, 2018

Reproduction is authorised provided the source is acknowledged.

1. Introduction

The European Commission is asking EFSA for scientific and technical support in the epidemiological analysis of animal disease outbreaks (i.e. African swine fever, lumpy skin disease and avian influenza) and to assess surveillance data (i.e. *Echinococcus multilocularis* and avian influenza). For this purpose, in the last years EFSA has been performing *ad hoc* data collections, gathering specific information on outbreaks, surveillance activities and concerned animal populations (i.e. poultry, domestic pigs, cattle and wildlife such as wild boar) related to these diseases (further named '*ad hoc* animal health data collections'). Data collections on animal populations are also implemented in relation to the EFSA-ECDC zoonoses summary report ([The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016](#)) and by EUROSTAT, but the resolution is not sufficient to be used in analytical epidemiology and risk assessments, mainly because the systems are designed for other purposes. In addition, EFSA is currently using *ad hoc* data models which are tailored to these diseases and to the related specific mandates, with a consequent lack of harmonisation across them and with the zoonoses data model.

At the moment, since EFSA data models are not completely harmonised; EFSA's data warehouse is not connected with the EC's Animal Disease Notification System (ADNS); and data submission and validation is not automated, the animal health data collections are labour intensive for both EFSA and the data providers of the Member States. Data transmission is now mainly done via e-mail exchange of Excel/xml files, which is time consuming, prone to copy/paste mistakes and does not exclude wrong reporting of variables. Therefore, there is a clear need to improve the technical aspects of the *ad hoc* animal health data collections to reduce the resource burden for both MSs and EFSA.

EFSA, therefore, decided to initiate a process of data model harmonisation and technical development on data submission, validation, analysis and reporting under the name of "SIGMA project", across the different data collection activities which are focused on the animal populations (i) affected by disease outbreaks or (ii) subject to disease surveillance.

EFSA's aim is to work together with Member States on the technical aspects of '*ad hoc* animal health data collections', in order to:

- Reduce the manual input of the data to be submitted by the Member States to EFSA
- Avoid double reporting to EFSA and, possibly, to other systems
- Provide the Member States with instruments to produce automatically draft national reports on animal health and surveillance in a protected environment (secure connections, log in credentials) to ensure data protection
- Increase the quality and the comparability of the data received from the Member States
- Shorten the time to retrieve up-to-date data, relevant for risk assessment purposes, and to release sound scientific advice

This report aims at: (i) providing the reader with the main elements to understand the SIGMA project and (ii) briefly describing the contents of the Animal Health Network meeting and at capturing the feedbacks having a direct impact on the strategy.

2. The SIGMA project in a nutshell

SIGMA has been planned as a three years' project with three main phases as described in the following sections.

Considering the complexity of the project and the ambition of creating a framework that could be used by all MSs, EFSA launched a call asking for support in the technical implementation of the project. The awarded consortium is led by the Istituto Zooprofilattico Sperimentale (IZS) Abruzzo e Molise "G. Caporale" and in partnership with the Friedrich Loeffler Institut (FLI), the Swedish National Veterinary Institute (SVA), the Bulgarian Food Safety Agency and the Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences.

The SIGMA consortium will provide technical support to interested MSs:

- to improve animal health data flows within the country (from the national data sources to the country data collection point –CDCP-, typically the national risk assessment body)
- to improve animal health data flows from the country data collection point to EFSA and
- to connect (preferably existing) tools for data analysis to the EFSA Data Warehouse to facilitate harmonised reporting by national and European risk assessment bodies.

2.1. SIGMA Phase 1

In this first phase, the main goals are:

- to design a harmonised data model, the SIGMA Animal Disease Data Model (σ -ADM) able to gather relevant data to address the requests related to the ongoing mandates (African Swine Fever, Avian Influenza, Lumpy Skin Disease, *Echinococcus multilocularis*) from the concerned MSs and from the existing data collection systems (ADNS/ADIS, WAHIS)
- to draw a comprehensive overview, at MS level (country card), of the institutions responsible for the collection of the data related to animal health and animal population (at this point in time, poultry, bovines and pigs)
- to outline the data flow within each MS and from each MS to EFSA to highlight potential drawbacks and propose technical solution to improve the system
- to provide a list of online tools for the data analysis and for the reporting of disease outbreaks / surveillance activities with the aim of make them available to the MSs to query the EFSA Data Warehouse, where the transmitted data are stored.

2.2. SIGMA Phase 2

The second phase will be mainly dedicated to the concrete implementation of the framework with the MSs that volunteer to take part to the pilot. This phase will be targeted on Avian Influenza (AI) and African swine fever (ASF). In detail:

- Support the volunteering MSs in designing Extract Transform Load (ETL) processes to select, transform and transmit the relevant national data in line with the σ -ADM
- Avian Influenza surveillance reporting: support on the transition from the EC system to the EFSA data collection framework (DCF)
- African Swine Fever: optimisation of the outbreak data collection
- Pig population: support on the setting of the data flows within and outside the volunteering MSs
- Analytical interactive online tools: connection of relevant online tools to EFSA's DWH to support ASF and AI disease outbreak analysis and reporting

2.3. SIGMA Phase 3

In the third phase, based on the lessons learned from the previous steps, the implementation of the SIGMA approach will be finalised and extended to the other relevant animal populations and MSs' activities. In particular:

- ETL processes: implementation of the Extract Transform Load processes (to select, transform and transmit the relevant national data in line with the σ -ADM) in the volunteering MSs
- Analytical interactive online tools: connection of relevant online tools to EFSA's DWH to support LSD disease outbreak analysis and reporting
- Poultry and cattle populations: support on the setting of the data flows within and outside the volunteering MSs
- Analytical interactive online tools: connection of relevant online tools to EFSA's DWH to support Avian Influenza and *Echinococcus multilocularis* surveillance analysis and reporting

3. The SIGMA - ADM

The main steps of the development of the σ -ADM are described in the following bullet points (see also Figure 1).

As a preliminary step, EFSA performed a **user case analysis** to identify all drawbacks and pitfalls related to the ongoing data collections in view of the related risk assessment and data analysis

- 1) The first step towards the harmonisation of the data collection was to **analyse and summarise the risk assessment requests** from the European Commission;
- 2) Once the risk assessment questions were retrieved, they were grouped into **categories**, each characterised by a common possible statistical approach. The outcome of this step was a set of **Envisaged Analysis** (i.e. a set of hypothetical statistical approaches that could address the identified risk assessment requests), essential to identify the data needs;
- 3) As each statistical approach needs **specific input data**, the Envisaged Analysis played a crucial role to identify and define in a concrete way the type of **data needed**
- 4) The data-need was then formalised in a data model: the **SIGMA Animal Data Model** (σ -ADM)
- 5) The σ -ADM was then **tested against all the ongoing *ad hoc* EFSA animal health data collections** to make sure that the experience gained over the last years was well integrated and the σ -ADM was comprehensive enough to encompass all type of risk assessment needs.

The first version of the model was then circulated among the members of the Animal Health Network and of the Animal Health and Welfare Panel to check its compatibility with all possible scenarios.



Figure 1: Sources of information in the process generating the SIGMA-Animal Disease Data Model (σ -ADM)

It is important to note that the data model resulting from this exercise will be tailored to fit the animal diseases for which EFSA has an ongoing mandate, i.e. African Swine Fever (ASF), Lumpy Skin Disease (LSD), Avian Influenza (AI), *Echinococcus multilocularis* (EM) and will be compatible with the zoonotic diseases included in the EFSA annual report (EFSA, 2017a). Other diseases will only be included if EFSA would receive a specific mandate on those to provide the risk managers with scientific information on those diseases.

The development of this scientific report is linked to the strategic objectives to widen EFSA's evidence base and optimise access to its data, build the EU's scientific assessment capacity and knowledge community and prepare for future risk assessment challenges (EFSA Strategy 2020, <https://www.efsa.europa.eu/sites/default/files/151008.pdf>).

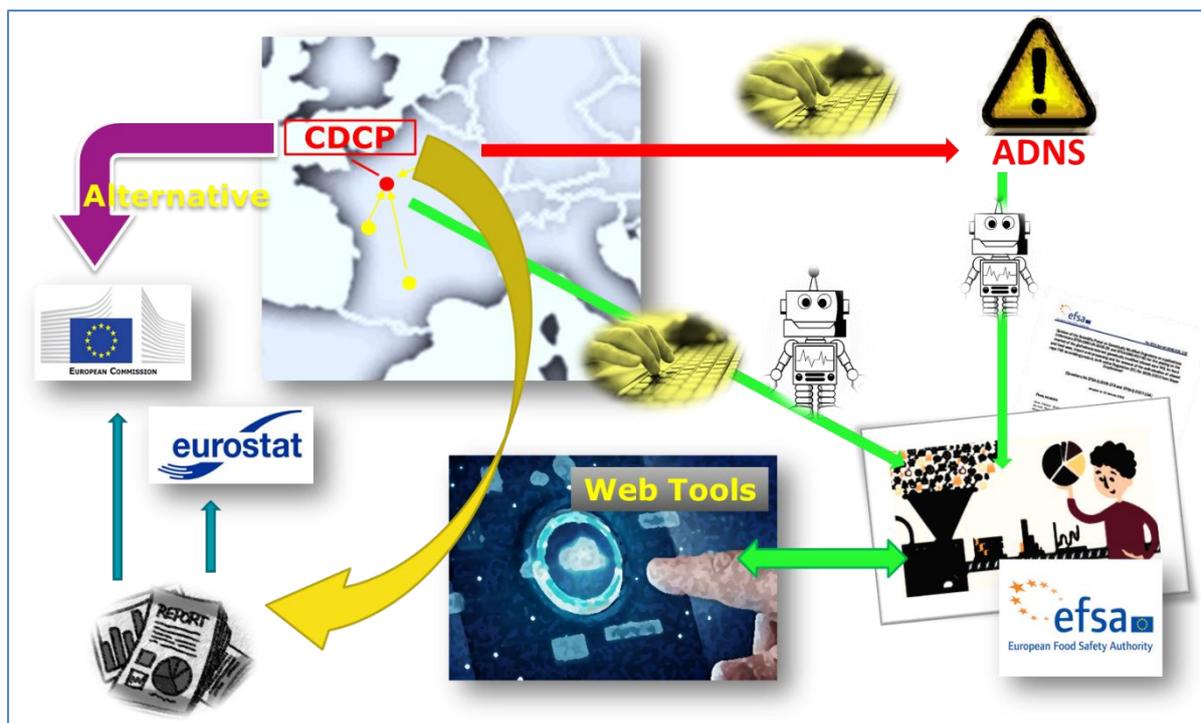
4. Network meeting report

The SIGMA project originates from a critical assessment of the current practices to collect data on animal population and animal diseases and has the purpose of optimising the entire process. For this reason, the project was outlined to the Network members in order to clearly define the scope and the objectives, and the related proposed strategy, with the explicit goal of seeking feedback on the suitability of the entire approach.

An important step in the process is represented by the development of a unique data model (σ -ADM), fitting the needs of animal disease risk assessment, across the different mandates received by EFSA, including the zoonotic agents. A set of entities (e.g. establishment), attributes (e.g. production type) and their possible values (e.g. breeding) were identified. Each element was given a definition following, where present, the wording used in the Animal Health Law. The data model was discussed in detail with the Network members and the comments reflected in the σ -ADM.

Based on the feedback received, the data flow has been reconsidered and the main features are described below (see also Figure 2):

- The animal population data (bovines, swine and poultry) could be collected at municipality level (NUTS 4/5)
- The MSs involved in the project, on a voluntary basis, identify/create a Country Data Collection Point (CDCP) where all relevant data converge from the national data sources
- The data from the national data sources will be transformed (by means of ETL processes – see Sections 2.2 and 2.3 for more information) to make them fit for the σ -ADM and will therefore be collected in the CDCP in a standardised format, ready to be transmitted to EFSA. This step avoids the MSs to change the definitions and the categories used in the national data collection systems in place: in fact, the ETL processes work as “translators” from a “data-language” to another.
- Once in the CDCP, the data can be checked by the MS and then transmitted to EFSA
- EFSA stores the data from all involved countries for risk assessment purposes as requested by mandates
- After submission/transmission of data, each MS will be able to access its own data, now harmonised at EU level, by means of secured connections (log in required) and use them for their needs
- The MSs can choose to use the interactive web tools, made available by EFSA and designed to produce graphs, diagrams, tables or full reports to fulfil the EU legislation, i.e. that have to be submitted to the EC by each EU country. As an alternative, they can still use their analytical tools and report directly to the relevant institution.



Yellow spots: National Data Sources; **CDCP:** Country Data Collection Point; **ADNS:** Animal Disease Notification System; **Keyboard icon:** manual input; **Robot icon:** automated process; **Web Tools:** web applications designed to produce the reports that the EC requires from the MSs; **Purple arrow "Alternative":** transmission of the reports that the EC requires from the MSs without using the web tools made available by EFSA

Figure 2: Data flow within the SIGMA framework

All proposed definitions for each item in the σ -ADM were discussed with the Network members and their input was captured in the new version of the Animal Disease data Model.

5. Conclusions and main messages

The main message from the Network which will have an impact on the SIGMA project and have to be reflected in the Scientific Report are:

- The SIGMA project is meant to be a service for the EU Member States and, as a consequence, the involvement is on a voluntary basis;
- The SIGMA project's aim is to support the development and implementation of technical features with the goal of reducing the burden for the MSs to submit *ad hoc* animal disease and population data to EFSA.
- The responsibility of the national risk assessment is entirely hold by the national authorities and EFSA will perform analysis only at European level (unless explicitly and formally requested by the EC and/or the concerned MS)
- The animal population data could be submitted to EFSA aggregated at municipality level (NUTS4/5), reducing the concerns about personal data confidentiality
- The frequency of data submission to EFSA will be agreed with the MS's according to the type of data and will be based on the outputs that are requested in the mandates.
- The involved MS should identify a Country Data Collection Point where all relevant data converge to be transformed and submitted to EFSA under full control of the MS. The Consortium awarded by EFSA could help in setting the national IT framework to make this data flow technically possible.