EXTERNAL SCIENTIFIC REPORT

APPROVED: 11 February 2022

doi:10.2903/sp.efsa.2022.EN-7254

Data submission from European countries to EFSA: the SIGMA project approach

SIGMA Consortium

Abstract

The Framework Working Contract (FWC) number OC/EFSA/ALPHA/2018/01 between the European Food Safety Authority (EFSA, the contracting authority), and the SIGMA consortium (the contractor), is in force from 23 May 2018 until 22 November 2021. The subject matter of the FWC is the provision of technical support to improve and automatize the collection and reporting of data on animal disease outbreaks and surveillance to EFSA by the European Union (EU) Member States and pre-accession countries. The EFSA project is named SIGMA. The contracting authority, among other tasks, asked the contractor the preparation of a Summary Report, describing the novelties and the benefits introduced by the implementation of the SIGMA framework.

Key words: Data flows, data mapping, data harmonization, data standardisation, animal health, animal population

Question number: EFSA-Q-2018-00054

Correspondence: BIOHAW@efsa.europa.eu

Suggested citation: SIGMA Consortium, 2022. Data submission from European countries to EFSA: the SIGMA project approach. EFSA supporting publication 2022:EN-7254 17 pp. doi:10.2903/sp.efsa.2022. EN-7254

ISSN: 2397-8325

Summary

The European Food Safety Authority (EFSA) has been receiving requests for support by the European Commission (EC) in the analysis of certain animal disease outbreaks and related risk factors (i.e. African swine fever, ASF; lumpy skin disease, LSD; avian influenza, AI). These types of epidemiological analysis require the results of surveillance and control activities, such as laboratory testing or vaccination actions, carried out by the European Union (EU) Member States (MSs) and pre-accession countries (IPA), and also the availability of high resolution and up to date data about the susceptible animal population, e.g. number and spatial distribution of establishments breeding susceptible hosts.

SIGMA Consortium has been appointed by EFSA to study and define the possible data collection flows from MSs and IPA to EFSA, considering all possible difficulties and related technical solutions to minimise the workload of the data providers, but assuring the level of data quality needed for risk assessment studies.

The approach proposed is based on the implementation of two steps:

a) A first step is aiming at reconstructing the syntax of the data flow with the minimum impact for the data providers. National repositories are created, where the data of interest are copied or exported by each country, according to the structure of their databases and their rules.

b) A mapping service (called SIGMA-EST: Extract and STandardise tool) takes care of recognising the codes used by the data providers and translating them into the standard EFSA's dictionary.

A specific technical roadmap was defined to identify all possible problems for the data providers and to support and to accompany them along the data collection process.

Firstly, EFSA produced a comprehensive overview of the competent authorities in the framework of the animal health and welfare, describing the ownership of the data domains and all the competent authorities in each MS and IPA involved in the generation, collection and storing of data on animal diseases. The results are elaborated in the format of Country Cards on Data sources on animal diseases and published in a public website.

Secondly, four type of questionnaires were prepared by the SIGMA Consortium (regarding pig and poultry population data, ASF and AI surveillance data) and submitted to all EU MSs and 4 IPA countries, including questions about the institutions generating or collecting the basic data, the level of details of collected data, the existing data repositories at local or national levels, the existing data exchanges between Institutions and technical means used.

Overall, 50 filled questionnaires were received from 21 countries (17 EU Member States and 4 IPA countries).

The following step was the preparation of the two data models: one for the population data and a second for the laboratory (surveillance) data. A simple tool, in Excel format, the SIGMA Digital Data Inventory (DDI) has been prepared to facilitate the preliminary analysis by the data providers of the mapping process between the national databases and EFSA data models.

Finally, a web-based data mapping tool, called SIGMA-EST (Extract and STandardise), has been developed. After the initial set up (mapping between national standards and EFSA standards) the tool is able to automatically "translate" the attributes and values as recorded in the national databases into the EFSA standards, according to the data models.

The SIGMA-EST tool provides a guided process to map the data in an automated way, creating a schema (Excel or in XML) for the export of national data.

Table of contents

1. Introduction

1.1. Background and Terms of Reference as provided by the requestor

This contract was awarded by EFSA to:

Contractor:

Leading Partner:
Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale" (IZSAM)
Campo Boario,
64100 Teramo, Italy
VAT registration number: 00060330677

Appointed as the leader of the group by the members of the group that submitted the joint tender and

Partner 2:
Friedrich-Loeffler-Institut, Bundesforschungsinstitut für Tiergesundheit (FLI)
Südufer 10, 17493 Greifswald
Insel Riems, Germany
VAT registration number: DE811354798

Partner 3:
Statens veterinärmedicinska anstalt (SVA)
Ulls väg 2B,
751 89 Uppsala, Sweden
VAT registration number: SE202100186801

Partner 4:
Bulgarian Food Safety Agency (BFSA)
15 A Pencho Slaveikov blvd,
1606 Sofia, Bulgaria
VAT registration number: PIC 959622359

Partner 5:
Estonian University of Life Sciences (EMU)
Fr.R.Kreutzwaldi 1,
51014 Tartu, Estonia
VAT registration number: EE100018015

collectively 'the contractor',

Contract title: Technical support to improve and automatize data collection and reporting on animal disease outbreaks and surveillance (SIGMA)

Contract number: OC/EFSA/ALPHA/2018/01

1.1.1. Background

The European Food Safety Authority (EFSA) has been receiving requests for support by the European Commission (EC) in the analysis of certain animal disease outbreaks and related risk factors (i.e. African swine fever, ASF; lumpy skin disease, LSD; avian influenza, AI). These types of epidemiological analysis require the results of surveillance and control activities, such as laboratory testing or vaccination actions, carried out by the European Union (EU) Member States (MSs) and pre-accession countries (IPA), and also the availability of high resolution and up to date data about the susceptible animal population, e.g. number and spatial distribution of establishments breeding susceptible hosts.

In the remit of the Framework Working Contract (FWC) number OC/EFSA/ALPHA/2018/01, the SIGMA Consortium has been appointed by EFSA to study and define the possible data collection flows from data providers to EFSA, considering all possible difficulties and relative technical solutions to minimise the workload of the data providers, ensuring, at the same time, the level of data quality needed for risk assessment studies.

The main objective of the service, therefore, was to provide EFSA with all necessary technical support in developing and implementing an automated data submission process to EFSA on:

- susceptible livestock,
- animal testing to detect animal diseases (LSD, AI, ASF).

To properly achieve this objective an innovative approach on the data collection framework was defined and various tools were developed for this purpose.

This report describes the data collection approach and the tools developed under the SIGMA activities as well as the main benefits introduced by the implementation of the SIGMA framework in comparison with the current and previous data collection approaches followed by EFSA.

1.1.2. Terms of reference

In the context of the FWC, the specific contract n.12 foresaw the preparation of a Summary Report describing the novelties and the benefits introduced by the implementation of the SIGMA framework.

In particular, the document should describe pros and cons introduced by the SIGMA framework by means of a comparison between "before" and "after" SIGMA. This comparison shall be performed on all aspects and all levels:

- The Country Cards.
- The data flows.
- The data models.
- The call for data and the data providers.
- The SIGMA Digital Data Inventory (DDI).
- The SIGMA EST (Extract and STandardise) tool.

1.2. Data collection processes at EFSA: a critical analysis

Currently, different approaches are in use by EFSA for the collection of data on animal health from the EU MSs and neighbouring countries.

For example, the collection of data on **LSD** was performed in many instances through Excel files, only partly in line with EFSA standards, which required significant additional efforts by EFSA personnel for the harmonisation of the data received. In addition, the data collected in this format could not be stored in a data-warehouse and the consistency over time was also jeopardised.

The data collection on **ASF** was performed through the EFSA Data Collection Framework (DCF), but with data models that were diverging from the standards in the Standard Sample Description ver. 2 (SSD2[1]). In fact, the SSD2 was not providing the full set of standards for the metadata that were necessary for the collection of laboratory data on animal samples. For this reason, the existing standards were somewhat enriched with additional variables and controlled terminology. However, such a deviation, made impossible to transfer the data from the DCF to the EFSA Scientific Data WareHouse (S-DWH).

In 2017, the "**Avian Influenza** overview" for the reporting period October 2016 – August 2017 (EFSA and ECDC, 2017[2]) was produced based, among other sources, on data submitted via email by the MSs as Excel files. More recently, the collection of surveillance data on AI was fully standardised, with the publication of a specific guideline for data reporting[3]. The SSD2 approach was followed for this data collection, asking the MSs to provide .XML files into the EFSA DCF, in compliance with the structure and values foreseen in the guideline and in the SSD2 catalogues. However, while ASF data were collected at test-result level, the AI data are still collected aggregated at NUTS3 level, hampering a full interoperability of the data with other information.

The approaches described above have limitations and shortcomings that must be considered.

The approach adopted for LSD implies a considerable extra effort by EFSA personnel and does not assure the comparability and harmonisation of data provided by the different countries. The approach adopted for ASF and AI, on the other hand, ensures a higher level of standardisation, but forces the data providers to adapt their own information systems to fulfil EFSA requirements. This entails a significant workload for the countries and the risk that important and critical data cannot be provided, because not available or not "exportable" according to the SSD2 standards.

This top-down approach, which is very common in the domain of data collection, implies theoretically less efforts by EFSA personnel in terms of data verification and manipulation. In reality, this approach has its drawbacks: one of these is the frequent rejection of a not negligible amount of data, due to the failure to comply with the SSD2 standards and its business rules. The difficulties of the countries to provide all required data according to SSD2 standards, may derive from various reasons, for example:

- Collection and registration of original raw data are not following the SSD2 standards and rules, forcing the countries to manipulate and modify *ex-post* the recorded data in the effort of complying with SSD2 requirements,
- Difficulties in retrieving data from national systems and transforming them according to SSD2 rules, due to limited technical knowledge by the services of national authorities,
- Difficulties in consolidating national data originally stored in different systems and/or different competent authorities.

The potential issues described above pertain to the **data standardisation**. Another problem that must be considered is the **data harmonisation**. i.e. the correct interpretation of the terminology used in the standards provided. A typical example is provided by the term "Backyard". Despite the apparently

[1] EFSA (European Food Safety Authority), 2013. Standard Sample Description ver. 2.0. EFSA Journal 2013;11(10):3424, 114 pp., doi:10.2903/j.efsa.2013.3424

[2] European Food Safety Authority, European Centre for Disease Prevention and Control, European Union Reference Laboratory for Avian influenza, Brown I, Mulatti P, Smietanka K, Staubach C, Willeberg P, Adlhoch C, Candiani D, Fabris C, Zancanaro G, Morgado J and Verdonck F, 2017. Scientific report on the avian influenza overview October 2016–August 2017. EFSA Journal 2017;15(10):5018, 101 pp. https://doi.org/10.2903/j.efsa.2017.5018

[3] EFSA (European Food Safety Authority), Aznar I, Bocca V, Gibin D, Papanikolaou A and Stoicescu A-V, 2020. Guidance for reporting 2020 avian influenza data. EFSA supporting publication 2020:EN-1867. 28 pp. doi:10.2903/sp.efsa.2020.EN-1867

clear meaning of the term (i.e. a relatively small place in which a relatively small number of animals is bred, mainly for own consumption), in reality the exact definition of this type of breeding varies across the different countries in terms of maximum allowed number of animals, obligation to register, commercial limitations, etc. As a consequence, the reporting of a farm as a "backyard" hides many potentially different entities with a different epidemiological role.

In addition, in both approaches followed by EFSA, data were requested according to a unique set of information (i.e. a unique data model), mixing population data (denominator) and laboratory/surveillance data (numerator). In most cases, however, the sources of these data are different (i.e. the authority for animal identification and registration for population data, and national/central veterinary laboratories for surveillance data) and therefore the national data providers are forced to collect the data at national level from the different sources, pre-manipulate and modify the data (which actually deviates from the official original ones), before submitting them into the DCF.

In summary, despite the general goal of having data available for addressing the mandates received by the EC was achieved, none of the two approaches can be considered fully efficient; both approaches are affected by possible serious deficiencies concerning the data quality and require some extra workload.

A different approach, more respectful of the data structures and the competencies at national levels ("ask the right thing to the right person"), is likely to be a better approach for collecting livestock data and laboratory data, which can be interlinked *a posteriori*, once in the S-DWH.

2. Methodologies

The new data collection flows from the EU countries to EFSA must have a minimal impact on the structure, organization and procedures already in place at national level. It must be able, also, to be adapted to the different national situations, in term of level of digitalization of data of concern, the distribution of databases where these data are stored and technical platforms in place.

The approach proposed is based on the implementation of a two-step data flow (**Figure 1**):

a) **Original national dataset transformation and loading.**
A first step is aiming at reconstructing the syntax of the data flow with the minimum impact for the data provides. National repositories are created, where the data of interest are stored by each country, according to the structure of their databases and their rules. The data providers do not have to prepare any standard file or to change the structure of the data as they have them stored. In addition, no re-codification work is asked. The sole condition is the presence of all required variables (mandatory attributes) in the copied/exported dataset. The frequency of this export will be defined according to official agreements between EFSA and the national competent authorities.

b) **Mapping.**
A mapping service (called SIGMA-EST: Extract and STandardise tool), after proper set up carried out by the data provider with the support of EFSA, takes care of recognising the codes used by the country and translating them into the standard EFSA's dictionary. Data providers have no longer to worry about the re-codification of its data message. SIGMA-EST serves as a "simultaneous translator", adapting the original file to the "language" of the recipients. It must be specified that the SIGMA-EST tool has its own catalogues of values, which in this case are aligned with those in use by EFSA, but the system could in theory manage different catalogues of values as reference and the following mapping with EFSA's values. The main added value of this service is the resolution of all mappings needed to interface with the external information systems. From an operational point of view, this functionality empowers the efficiency of the system, considering that at present each country spends a lot of time and resources to map the contents of its databases with several different external systems, including EFSA.

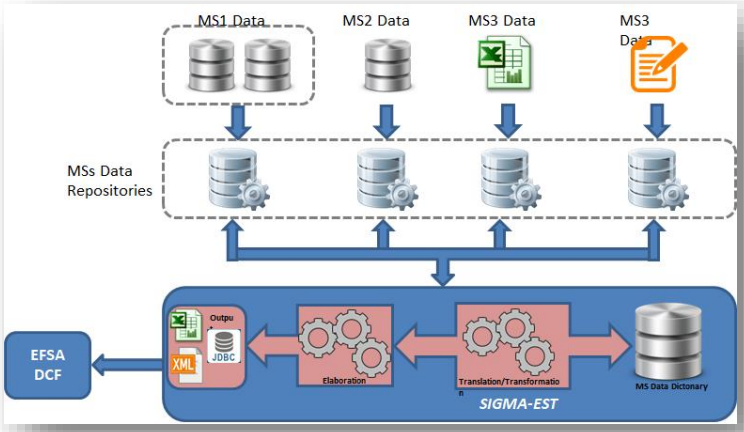

Figure 1: Conceptual framework of the data collection system from the EU MSs to EFSA DCF

However, although the proposed data collection approach seems to be more consistent and efficient than the previous / existing data flows in place at EFSA, changing the current and already established data collection approaches cannot be considered an easy process from the data providers point of view. Therefore, a specific technical roadmap was defined to identify all possible problems for the MSs and to support and to accompany them along the data collection process (**Figure 2**). Various tools were also developed (see next chapters).

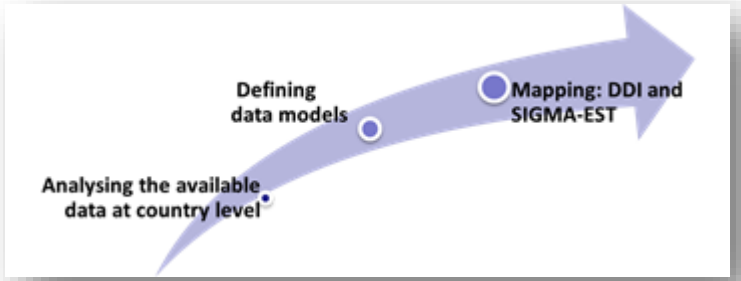

Figure 2: Roadmap for the data collection from the EU MSs to EFSA DCF

2.1. *Questionnaires on national data flows*

Considering the "bottom-up" approach introduced by the SIGMA project, the first step was the collection of all available information about the organization of data flows at national level.

For this purpose, EFSA produced a comprehensive overview of the competent authorities in the framework of the animal health and welfare, describing the ownership of the relevant data domains and

all the competent authorities in each country involved in the generation, collection and storing of data on animal diseases. The data are retrieved by means of an online survey. The results are elaborated in the format of Country Cards on Data sources on animal diseases and published in a public website[4].

In addition to this preliminary picture of the national Institutions involved in the collection of data on animal diseases, more specific information on the organization of the data flows in each country should be gathered. For this objective, the following questionnaires were prepared:

- For African swine fever:
    - Questionnaire regarding the technical description of the data repositories and the related data flows of **pig population data**;
    - Questionnaire regarding the technical description of the data repositories and the related data flows of **ASF surveillance** in pigs and wild boars.
- For Avian Influenza:
    - Questionnaire regarding the technical description of the data repositories and the related data flows of **poultry population data**;
    - Questionnaire regarding the technical description of the data repositories and the related data flows of **avian influenza surveillance**.

The questionnaires include questions about the institutions generating or collecting the basic data, the level of details of collected data, the existing data repositories at local or national levels, the existing data exchanges between Institutions and technical means used. The possibility of data exchange with EFSA is also explored.

The questionnaires were submitted to the EU MS, directly through the Consortium partners; through the EFSA Scientific Network for Risk Assessment (RA) in Animal Health and Welfare (AHAW) and through the EFSA Focal Point network. Also four IPA countries, namely Montenegro, Serbia, North Macedonia and Kosovo were involved.

Overall, 50 filled questionnaires were received from 21 countries (17 EU Member States and 4 IPA countries).

The understanding of the nature of the data (digital or not, level of resolution, management, etc.) allows EFSA to elaborate tailored support to easy the work of the data providers in submitting data to EFSA.

2.2. *Digital Data Inventory*

The following step was the preparation of the two data models: one for the population data and a second for the laboratory (surveillance) data.

The choice of preparing two distinct data models was in line with the principle of "asking the right data to the right persons". In fact, in several countries the Institutions or Authorities collecting and storing the data on animal populations are different from those dealing with laboratory data. The creation of two different data models, therefore, is coherent with the objective of avoiding asking additional efforts to the countries for the collection of all requested data.

The two data models are designed to collect data at the lowest level of aggregation:

- For population data, the establishment and subunit levels are considered (**Figure 3**). Individual animal level has been considered not necessary for EFSA purposes and it would have introduced additional unnecessary difficulties in maintaining the information up to date. Total number of animals in the subunit level at the time of data extraction are requested. The definitions of the

[4] https://www.efsa.europa.eu/en/search?s=Data+sources+on+animal+diseases%3A+Country+Card

Animal Health Law (AHL) are considered and new specific catalogues have been developed (or adapted to the already existing ones).

- For laboratory data, the sample level, with all results, is considered and based on SSD2 standards.

The link between population and laboratory data is assured by the **subUnitId** (Primary Key), which is the unique repeated data in the two data models.

Interoperability with the ADIS is assured by means of the ADIS notification number (if provided).

The availability of host population data is fundamental for any risk assessment. Indeed, these data are crucial for any type of statistical and epidemiological analysis (denominator). The collection by EFSA of population data from the EU MSs and IPA countries would result in the creation of a comprehensive database at EU level, currently not available in any other EU Institution. This database could be made available also to other EU Agencies and Institutions, in line with the relevant legislation in force on data protection and upon agreement with the data providers.

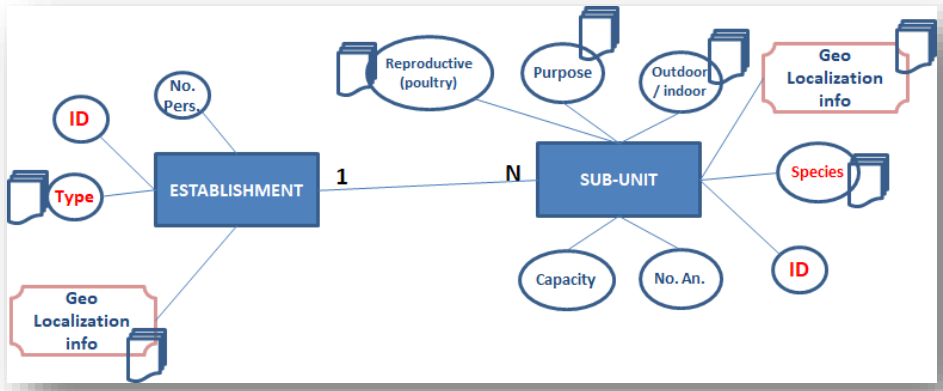

Figure 3: Diagram of the structure of population data model

The laboratory data model was based on SSD2 standards, but tailored and adapted to animal health needs, avoiding major redundancies in the list of attributes and catalogues' values. For this purpose, the revision of the existing SSD2 attributes, the definition of new ones and the review of the catalogues' values and definitions were made. Despite the simplification efforts and the reduction of data requested to only those essential for animal health, the resulting data model is rather complex, accounting for a list of 57 attributes grouped in 5 entities, as reviewed by the SIGMA project (**Figure 4**).

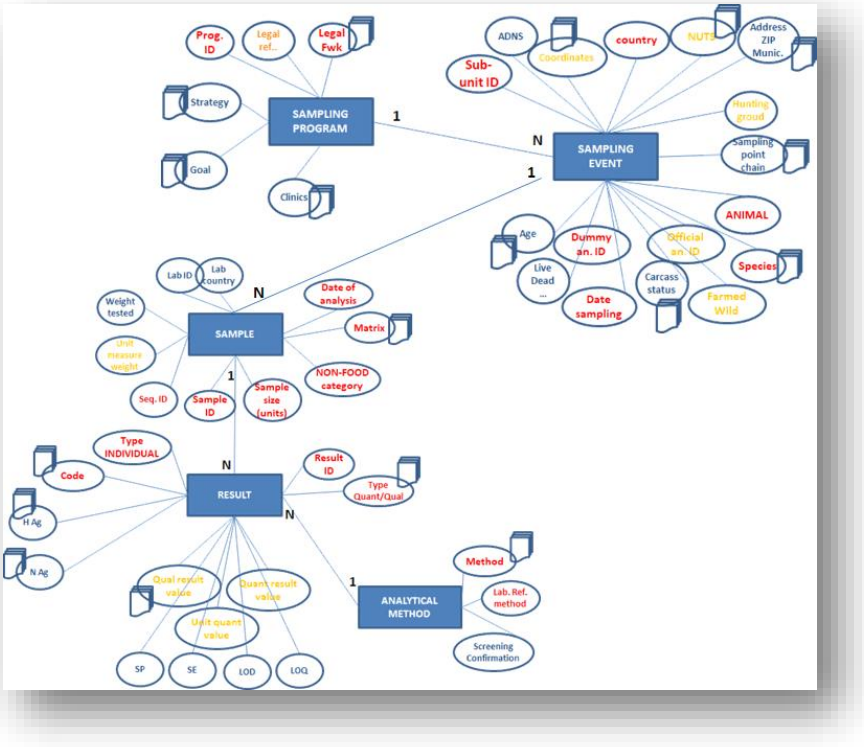

Figure 4: Diagram of the structure of laboratory data model

Considering the complexity of the data models, which implies some basic users' knowledge on the structure of a database and the meaning of "entities", "attributes", "catalogues", etc.., to facilitate the mapping process between the national databases and EFSA data models by the data providers, the SIGMA Digital Data Inventory (DDI) for population and laboratory data models have been created (**Figure 5**).

DDI is a simple tool, in Excel format, which is familiar to the great majority of people (even with limited skills in informatics). DDI guides the user through the attributes and catalogues' values, allowing him/her to indicate whether that specific attribute and catalogue's value is present in the national database and giving the opportunity to write the correspondent national name of each attribute / variable.

The use of this DDI tool by data providers is propaedeutic to the final mapping exercise in the SIGMA-EST tool.

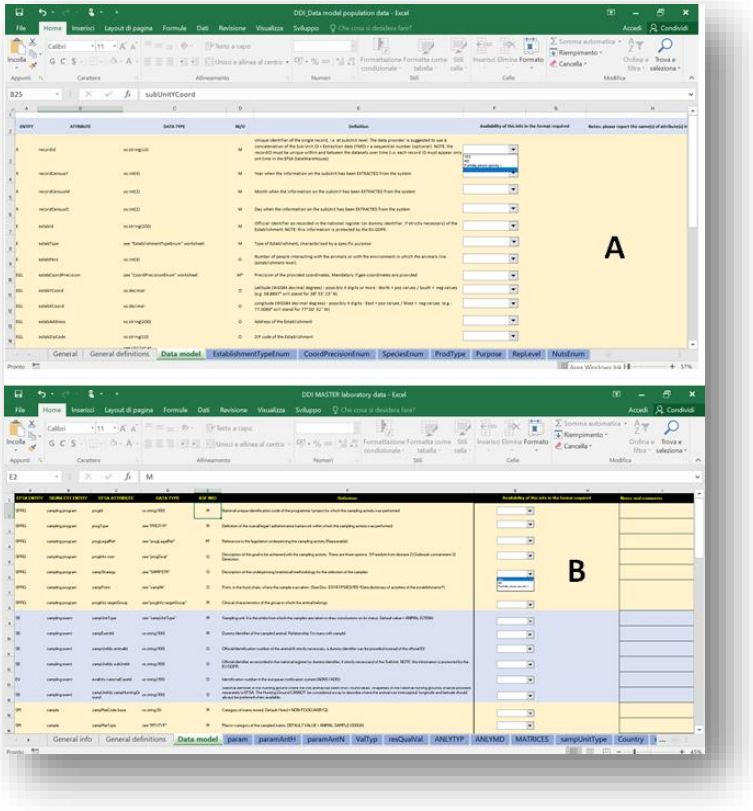

Figure 5: SIGMA Digital Data Inventory (DDI) for population (A) and laboratory (B) data models

2.3. *SIGMA-EST web mapping tool*

The SIGMA-EST (Extract and STandardise) tool, is the "translator" able to automatically "translate" the attributes and values as recorded in the national databases into the EFSA standards, according to the data models.

The SIGMA-EST tool provides a guided process to map the data in an automated way, creating a schema (Excel or in XML) for the export of national data. Once the mapping has been set, the data provider simply extracts the relevant data from its national database and upload them in the SIGMA-EST which translate them into "EFSA standard language" and produces the XML file to be submitted in the DCF.

The main aim of this approach is to reconstruct the syntax of the data flow with the minimum impact for the data providers.

The SIGMA-EST is an "expert system", which analyses the country dataset, recognising the variables of interest and reconstructing the final standard data structure to be loaded into the EFSA DCF. At first stage, pre-validation data quality checks are performed on the respect of data congruity (aggregation levels, presence of mandatory data, etc.).

This approach ensures the maximum flexibility of the data collection system, allowing EFSA to decide the level of aggregation needed for each variable and also the possibility to upgrade the system,

including additional variables, given that these additional data are included in the original national dataset.

SIGMA-EST is a mapping service, able to recognise the codes/syntaxes used by the country and to translate them into the standard EFSA's dictionary. The detailed information on data dictionaries utilised at national levels are the basis for the preparation of the mapping system, which is based on "databases of mapping tables" for each data provider, where the precise definition of each variable of interest within the national database is set.

The SIGMA-EST system utilizes a "semantic" logic. In fact, the original message to be mapped can have different structures, but the system will recognise the "tags" of interest. The "translation" service is designed to be parametric and independent from the information that the two "actors" (the sender and the receiver) intend to exchange, and therefore it is possible for an operator to "automatically" add new mapping specimens without any modification to the software.

It is important to highlight that the SIGMA-EST system is reversible. The data providers can download the data submitted to EFSA and revert them automatically according to the codes in use at national level by means of a re-codification process embedded in the SIGMA-EST.

In addition, the same approach could be also considered for translating all tags and labels from one language to another, using I18N technique.

The SIGMA-EST system is structured according to (**Figure 6**):

a) Domains

b) Entities

c) Activities

A **Domain** (also called catalogue) is a set of values of a specific type. For example, the list of values used to indicate:

- farm's production type,
- the animal species,
- animal breed,
- animal categories according age (e.g., less than 1-year-old, 1-2 years old, etc.),
- animal categories according the production type (e.g., animals for milk production, animals for meat production, animals with multiple attitude, animals for egg production, etc.),
- the laboratory diagnostic methods, including possible hierarchies (e.g.: direct and indirect methods),
- type of tested samples (e.g., whole blood, serum, blood with anticoagulants, organs, tissues, etc.),
- etc…

An **Entity** is an object containing a set of properties. For example, the table containing data on:

- farms,
- individual animals,
- vaccinated animals,
- tested animals,
- tested samples,

- results of laboratory testing,
- etc…

The single properties of an Entity can refer to a Domain or not. For example, in the case of FARMS Entity, the property describing the animal species kept must refer to SPECIES Domain, whereas the number of animals kept is simply described by an integer numeric value.

An **Activity** is a combination of one or more entities. The entities involved in an activity may be related each other or not. For example, the data flow regarding animal vaccination includes the following Entities: FARMS, ANIMALS and VACCINATION. Specific rules must be defined on how the different entities are linked each other within the same activity flow.

Following this structure, the "translation" of country data into the EFSA dictionary is considering basically the two aspects: the **semantic** and the **syntax** of the message.

For example, within the data flow on animal vaccination for LSD, the semantic translation from the species code used by the country A (value: "0101") to the code used by EFSA (value: "A006581A") for "Cattle" has this kind of structure:

Activity(Vaccination),

Entity(Animal),

Domain(Species),

Value(Cattle), CodeCountryA(0101), CodeEFSA(A006581A)

Concerning the syntactic translation of the message, for example, in case in the original database of the country A, the information on animal breed ("Holstein Friesian", "Simmental", "Jersey", etc.) and the category ("bull", "heifer", "calf", etc.) are registered using a single property, the system can recognised the single values and split them into two new properties:

Activity(Vaccination),

Entity(Animal),

Property Country A(Breed_Category), Property1EFSA(Breed), Property2EFSA(Category)

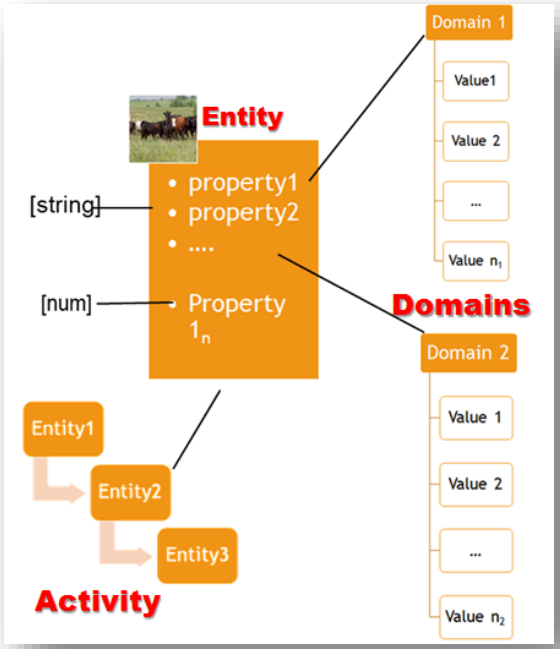

Figure 6: Conceptual structure of the SIGMA-EST system comprising Domains, Entities and Activities.

3. Conclusions

One of the main objectives of the SIGMA project is to provide the EU MSs and IPA countries with simple tools for "translating" the data stored in their national databases into a standardised dataset, harmonised across countries, usable for EFSA during the risk assessment process.

In fact, EU national authorities already spend a lot of time in adapting their information according to several different requirements foreseen in the framework of various data flows towards international Institutions. In addition, the amount of data digitally exchanged among the EU food safety and animal health national authorities is exponentially increasing, requiring a continuous and probably increasing effort on data manipulation and analysis by the countries providing data. Any solution, therefore, able to facilitate the exchange of new data, without determining additional workload for the national authorities, can be only seen under a favourable view.

At the moment of the publication of this report, the quality assessment of the ASF data from 2020, collected using the SIGMA framework is under evaluation. The first outcomes are promising and the full report will be published in 2022.

The present report aims at describing the technical aspects of the SIGMA approach, which is proposed as a model and suggested for other EFSA data collections on animal diseases surveillance and related population data, from the EU MSs, IPA countries and any other interested country. Moreover, considering the high flexibility of the approach and of the instruments made available, the approach is potentially suitable for any data collection.

The classical top-down approach followed in the design of data collection framework, especially at the supranational level, has been completely reversed in the SIGMA project. The first question was "Which kind of data the country can easily and reasonably provide?". This bottom-up approach, albeit

apparently more time demanding, allows to calibrate the data collection flow to what practically is gathered at national level and generated at local level. In fact, in non-synchronous and not real-time information systems, since the request for data is time (sometimes months) after the event (e.g. animal sampling), the possibility of *ex-post* intervention and integration of the original dataset collected at national level is very limited, often impossible. This is the main cause of poor data quality and incompleteness of data collection at supranational level.

Even the efforts made by national authorities to include in their information systems all the information and the specific requirements set by the International Institutions are often doomed to failure, given the impossibility to respect various, different and sometimes conflicting, external standards and data models.

Especially when the data are provided on voluntary basis, national authorities may be discouraged to make available their dataset, when a lot of work must be performed to adapt the data to the new standard.

The approach proposed in the SIGMA project, and described in this report, allows the harmonisation and collection of detailed data from the countries with limited efforts from the national authorities. Only a preliminary work for setting the mapping of attributes and values is requested, using the SIGMA-EST tool specifically developed for this purpose.

A not obvious added-value of SIGMA approach is also its scalability. In fact, the same solutions and tools, like the SIGMA-EST, useful for collecting data from the European countries, can be used to create or consolidate a national database. For example, it is rather common that, when more than one veterinary laboratory is involved in a diagnostic activity in a country, these laboratories register data on their activities by different Laboratory Information Management Systems (LIMS). Collection of harmonised detailed (sample-based) data from all the laboratories using different LIMS is often a big challenge, requiring extra-efforts in data mapping and trans-codification. SIGMA-EST may represent a useful tool also for the collection of diagnostic data from different laboratories in a country, avoiding to force them in changing their own LIMS and their data manipulation procedures.