



EXTERNAL SCIENTIFIC REPORT

APPROVED: 9 November 2022

doi:10.2903/sp.efsa.2022.EN-7667

Applying the Darwin Core data standard to wildlife disease – advancements toward a new data model

ENETWILD consortium¹, Francesca Jaroszynska², Guillaume Body², Sophie Pamerlon^{2,3}, Anne-Sophie Archambeau^{3,4}

²Office Français de la Biodiversité

³GBIF France (Global Biodiversity Information Facility), Paris, France

⁴IRD (Institut de recherche pour le développement), Paris, France

Abstract

Wildlife diseases have an impact on biodiversity, economy, and public health. As this impact increases with climate change, land-use change and trade, a multidisciplinary approach in analysing data from these different disciplines is increasingly required. The need for a common data standard to improve data sharing of surveillance efforts is thus larger than ever to enable transnational proactive and reactive measures to be taken. The Enetwild consortium, funded by the European Food and Safety Authority (EFSA), evaluated the capacity for the Darwin Core standard to handle wildlife disease data. In addition to two new terms previously proposed by Enetwild, the Darwin Core standard proved to have the flexibility to handle complex datasets from a wide variety of research and surveillance sources from local to international scales. Integration of the data model to the Global Biodiversity Information Facility is discussed, and controlled vocabularies specific to epidemiological data are proposed.

© European Food Safety Authority, 2022

Key words: Data standard, Darwin core, wildlife disease

Question number: EFSA-Q-2022-00055

Correspondence: biohaw@efsa.europa.eu

¹ www.enetwild.com

Disclaimer: The present document has been produced and adopted by the bodies identified above as author(s). This task has been carried out exclusively by the author(s) in the context of a contract between the European Food Safety Authority and the author(s), awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Acknowledgements: We would like to thank Alessandro Broglia and Lina Mur for their help with the SIGMA and ADIS databases, and general wildlife disease expertise.

Suggested citation: ENETWILD consortium, Francesca Jaroszynska, Guillaume Body, Sophie Pamerlon, Anne-Sophie Archambeau, 2022. Applying the Darwin Core data standard to wildlife disease – advancements toward a new data model. EFSA supporting publication 2022:EN-7667. 68 pp. doi:10.2903/sp.efsa.2022.EN-7667

ISSN: 2397-8325

© European Food Safety Authority, 2022

Summary

Wildlife diseases have an impact on biodiversity, economy, and public health. Better knowledge of disease patterns would be to the benefit of conservation measures, livestock production and, thus, ultimately human health. Nevertheless, disease surveillance systems that operate at an international scale have often incompatible data structures with other wildlife disease data and with wildlife distribution data, inhibiting effective data sharing and proper design of preventive measures, and thus rapid transnational responses to disease outbreak. As the impact of diseases on biodiversity, economy and public health increases with climate change, land-use change and trade, the necessity for a common data standard to improve data sharing of surveillance efforts is larger than ever to enable transnational proactive and reactive measures to be taken. Additionally, there is an interest in bringing epidemiological data closer to biodiversity data, thereby simplifying their management and compatibility.

To address these broad issues, the Enetwild consortium, funded by the European Food and Safety Authority (EFSA) was formed to collect existing data on wildlife host abundance and distribution in Europe. Alongside data on their associated pathogens, the consortium proposed a data model to aggregate the host data according to Darwin Core standards¹. The Darwin Core Standard is an international standard developed to enable biodiversity data sharing, consisting of terms with fixed definitions with as simple a structure as possible.

However, we do not know whether the complexity of wildlife disease data can be easily captured in the current version of the Darwin Core standard, often comprising complex data structures with species interactions. Firstly, wildlife disease data consist of observations of both the host and the pathogen, whereby each positive case of a disease is associated with the observation of the host species. Secondly, disease diagnosis is primarily conducted on a tissue sample extracted from the host, thus separating the in-field host observation from the laboratory-diagnosed disease occurrence. In order to capture these types of interactions and sub-setting, Darwin Core must be able to handle hierarchical data structuring.

Here we propose a data model for wildlife disease data following the Darwin Core standard. The wildlife disease data model is structured around an observation event, describing the process of observation in a particular location and point in time, i.e. the observation of a host species (wildlife host occurrence), from which a tissue sample is extracted and tested, leading to the disease identification (disease occurrence). We propose nesting in the Occurrence table to handle this hierarchical structuring of the host and the tissue sample, whereby an Occurrence observation of the host can be parent to a sub-occurrence of a tissue sample taken for laboratory analysis, e.g. a wild boar observation and a collected blood sample, respectively. To differentiate between these two types of information, one an observation of an animal host and the other a tissue sample, we use the *basisOfRecord* term to capture this. Thus, an observation of the host is a *human observation* and a tissue sample a *material sample*. Hierarchical structuring relies on the implementation of the term *parentOccurrenceID*, whereby each 'parent' occurrence of the host species may be associated with multiple 'child' occurrences or material samples. Here we propose a new use case for the *parentOccurrenceID* term, which was previously demonstrated to be able to handle sub-occurrences containing details about parts of a parent group observation (such as life stage or gender).

Using the Darwin Core standard, we propose a data model that would allow effective harmonisation of wildlife disease data. Finding solutions to capturing complex hierarchical biotic interactions in Darwin Core is currently underway at GBIF, though relying on a separate relationship table. However, we advocate for the introduction of the new *parentOccurrenceID* thanks to its simplicity and very general applicability, being adaptable to any group occurrence where detailed or partial information is available, or to hierarchical interaction relationships such as between hosts and pathogens. We believe the employment of the new term would be complementary to the current GBIF developments, and of benefit to many Darwin Core users where details apply to differing levels of hierarchical occurrences.

Table of contents

Summary	3
Table of contents	4
1. Introduction	5
1.1. Background and Terms of Reference as provided by the requestor	5
1.2. Scope of the report	5
1.3. Darwin Core	5
1.4. Challenges and problem framing	7
1.4.1. Capturing the host-pathogen relationship	8
Proposition	9
1.4.2. Handling multiple laboratory analyses from one individual	11
Proposition	11
1.4.3. Handling multiple individuals in one laboratory analysis	14
Proposition	14
1.4.4. Coping with virus strains and variants and their identification	15
Proposition	16
1.4.5. Host data with details at the group level	17
1.4.6. Avoiding duplication of dates	18
Proposition	18
1.4.7. Statistical data	19
Proposition	19
1.4.8. Methodological and Sampling Information	20
2. Formatting propositions for epidemiological data	21
3. Case studies	23
3.1. CSF/ASF EU database	23
3.2. Tuberculosis in UK badgers	27
3.3. SIGMA database	31
3.5. Preparation for Darwin Core developments	43
4. Conclusions and recommendations	44

1. Introduction

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

Contract OC/EFSA/ALPHA/2016/01 - 01 was awarded to the Universidad de Castilla-La Mancha by the European Food Safety Authority (EFSA) under the title: Wildlife: collecting and sharing data on wildlife populations, transmitting animal and human disease agents. The terms of reference relevant to this task under sub-contract 9 were to develop standards for epidemiological data for key wildlife diseases in Europe.

1.2. Scope of the report

Wildlife diseases have an impact on biodiversity, the economy, and public health. There is currently an increase in infection and transmission risk with climate² and land use change combined with an increase in the population density of certain host species. Better knowledge of disease patterns would be to the benefit of conservation measures, livestock production and ultimately human health.

In order to gain knowledge on disease circulation in Europe and react faster to disease outbreaks, there are multiple surveillance systems in place around Europe. These surveillance systems operate from sub-national to international scales. However, data are generally stored in non-harmonised formats, inhibiting uptake of reactive and proactive measures to disease outbreak. Furthermore, the few existing standards in place for wildlife or animal disease data do not correspond to the data standards for wildlife distribution data, additionally preventing a liaison between host and pathogen distributions. National initiatives result in local response to disease, however given the numerous land borders and trade routes amongst European member states a European-wide perspective on the state of disease surveillance is needed.

In order to improve disease surveillance in Europe, better data- and data storage standards need to be developed that facilitate European-wide aggregation and harmonisation of information on wildlife distribution and their associated pathogens. An initiative was launched by the European Food Standards Authority in 2016 (EFSA) to aggregate data on the abundance and structure of wildlife host populations and their pathogens. This initiative, in the form of a European consortium of wildlife and epidemiology experts (Enetwild) first began by developing a data standard for population and abundance data for selected wildlife host species (Enetwild consortium et al. 2020).

Building upon this initial phase, the second phase involves a similar process of data standard development for disease data. Until now, no such standard has been achieved using Darwin Core, but if developed it would enable data compatibility with data on host species abundance and distribution. Here we explore the capacities of the Darwin Core (DwC) standard for harmonising epidemiological data, and propose a standard building on the latest Enetwild data model and recent developments of the Darwin Core standard.

1.3. Darwin Core

Darwin Core is an international standard developed to enable biodiversity data sharing³. Initially developed in 1999, and formalised by the Darwin Core Task Group in 2009, the standard maintains a list of terms with fixed definitions with as simple a structure as possible. The standard has seen growing usage, particularly for protocolled data (EU BON), ocean observations (OBIS) and global biodiversity observations (GBIF). Record types such as herbarium data have been described by Darwin Core (DwC) since its launch, and recent developments now allow treatment of geological data in a subsection dedicated to add more context to paleontological data. Camera trap data has also been incorporated

into the Standard, opening many possibilities for wildlife surveillance data storage⁴. New uses of DwC will be implemented in the next GBIF data model to facilitate the description of, for example, genomic data, sampling data, traits, and environmental information.

In its simplest form, Darwin Core consists of a flat data format, consisting of one data table (Simple Darwin Core). For example, Tables 1a-1b demonstrate the conversion of simple species occurrence data to DwC. This simplicity maintains human and machine readability as much as possible. See Wieczorek et al.³ and Groom et al.⁵ for further details on the Darwin Core configuration.

Table 1a: Simple occurrence data as recorded in the producer format (original variable name, date and language)

species	nr observed	location	Date
wild boar	1	Auriol	19/03/2020
wild boar	1	Auriol	22/03/2020
wild boar	1	Cabriès	30/01/2018

Table 1b: Simple Occurrence data translated to the Darwin core standard. The *basisOfRecord* denotes the specific nature of the record, and *individualCount* the number of individuals present at the time of the Occurrence.

occurrenceID	basisOfRecord	date	locality	scientific Name	vernacular Name	individual Count
occurrence1	human observation	2020-03-19	Auriol	Sus scrofa	wild boar	1
occurrence2	human observation	2020-03-22	Auriol	Sus scrofa	wild boar	1
occurrence3	human observation	2018-01-30	Cabriès	Sus scrofa	wild boar	1

The simplest flat format structure is often inadequate for certain data sets, and in most cases a star schema is employed to accommodate more complex data structures whereby a record contains a central event, Occurrence or Taxon core and, where necessary, multiple extensions. This relational set-up is known as the Star schema. Relational tables have been introduced and continue to be introduced to accommodate these complex scenarios, some being applicable to more commonly reported information (such as Location information), though often discipline-specific (as indicated in Table 2). Additionally, information can be stored in tables associated to other relational tables, creating indirect linkages with the Event core, such as measurements corresponding to an Event, or to Occurrences associated to a particular Event (Figure 1).

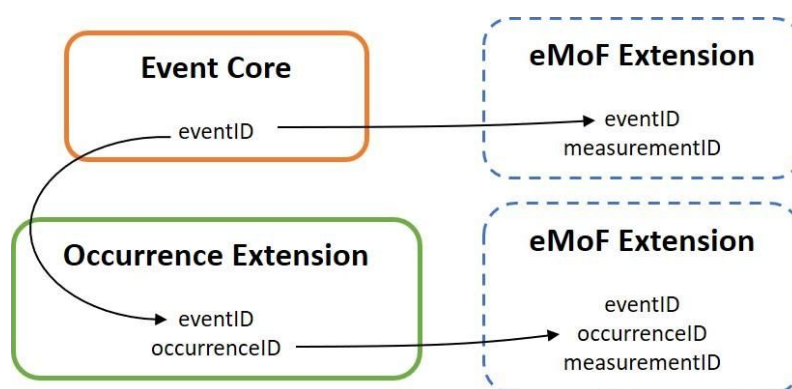


Figure 1: conceptual diagram of the association between the Event Core and the Occurrence and extended MeasurementsOrFacts (eMoF) extension as introduced by OBIS⁶.

1.4. Challenges and problem framing

Despite the fact that the Darwin Core standard has been embraced by multiple disciplines, such as marine ecology and palaeontology, we still lack development for wildlife disease and surveillance data. Indeed, we do not know whether the complexity of wildlife disease data is easily captured in the current version of the Darwin Core standard, since the data often comprise complex data structures including host and pathogen occurrences, tissue sampling, and species identification from laboratory analyses.

Recent developments via the Enetwild project have resulted in the capacity for Darwin Core to handle complex data on the distribution and abundance of multiple wildlife species, raising hopes that complex epidemiological data sets will also be able to undergo standardisation. A preliminary data model was first developed for the wild boar (*Sus scrofa*), and later expanded to incorporate ungulate and carnivore species data in Europe¹.

The next step towards the project's overall objective to link disease data with wildlife data is to determine whether disease data can be standardised within the Darwin Core Framework, and if so, whether developments to the current Darwin Core model structure need to be adapted. Below, we outline the major challenges to accommodating epidemiological data within the Darwin Core framework. These fall into the following themes:

- handling the relationship between the host and pathogen
- handling multiple laboratory analyses extracted from one individual
- multiple individuals pooled into one sample for laboratory analysis
- storing information on virus strains
- storing different dates corresponding to different stops in the observation and analysis process
- storing information on sampling procedures and protocols.

Each of these components can be simplified into when and where an observation was made, what the observation consisted of, and whether measurements were taken either of the observation or of the event. These are all features that, in theory, can be mapped to existing Darwin Core relational and core tables. The following Darwin Core classes were deemed necessary for producing the Epidemiology data model:

Table 2: An overview of the Darwin Core classes and extensions used in the development of the proposed data model

Class	Description	URL
Record-level	terms that are generic in that they might apply to any type of record in a dataset	
Event	An action that occurs at some location during some time	http://rs.tdwg.org/dwc/terms/Event
Location	A spatial region or named place	http://purl.org/dc/terms/Location
Occurrence	An existence of an Organism (sensu http://rs.tdwg.org/dwc/terms/Organism) at a particular place at a particular time	http://rs.tdwg.org/dwc/terms/Occurrence
Taxon	A group of organisms (sensu http://purl.obolibrary.org/obo/OBI_0100026) considered by taxonomists to form a homogeneous unit	http://rs.tdwg.org/dwc/terms/Taxon
MaterialSample	A physical result of a sampling (or subsampling) event. In biological collections, the material sample is typically collected, and either preserved or destructively processed.	http://rs.tdwg.org/dwc/terms/MaterialSample
Identification	A taxonomic determination (e.g., the assignment to a taxon).	https://dwc.tdwg.org/terms/#identification
extended measurementOrFact	A measurement of or fact about an rdfs:Resource (http://www.w3.org/2000/01/rdf-schema#Resource), extended version linking to occurrences	https://rs.gbif.org/extension/obis/extended_measurement_or_fact.xml

Hierarchically structured data are common throughout wildlife surveillance data. Firstly, during the Event core we find transect-point hierarchy, whereby multiple points can be sampled along a transect. Secondly in the Occurrence extension we find hierarchy in cases where partial information is known about an observation of a group of individuals (e.g. 7 observed individuals wherein 3 adult females, 2 juveniles), to cases of partial identification (some bird rings identified in a flock of birds). Finally, hierarchy can be found in measurements, for example when an estimation is associated with its confidence intervals. Epidemiological data are structurally not dissimilar, wherein tissue samples collected for disease identification can be subsampled, or a subsample of a collection of host individuals is selected for testing.

Recent demands have been made to incorporate hierarchical structuring of datasets into Darwin Core, firstly in the Darwin Core Event Core, whereby events may happen within events. A second demand was made by the Ocean Biodiversity Information System (OBIS) for the MeasurementsOrFacts table, whereby events may have occurrences, and occurrences may have measurements. This has resulted in the development of the *extended* MeasurementsOrFacts extension. Finally, a demand was made in the previous iteration of the Enetwild wildlife data model¹ to allow measurements of measurements, such as the uncertainty around an estimated population density. However, wider adoption of hierarchical data structures in Darwin Core remains limited (although see Enetwild Consortium et al.¹).

The premise of zoonotic disease observations is the presence of a host observed at a particular location at a particular time and the identification of a pathogen. The location and surveillance timing correspond to the Event core. The presence of the host and the pathogen are further described in the occurrence extension. However, how we define the occurrences, and how these are arranged and liaised, will determine the structuring and linkages between the measurements, the results, and the observations. We discuss these points below in light of the recent Darwin Core developments. We use existing Darwin Core terms and terms under proposition in the previous Enetwild data model wherever possible to limit an inflation of term dictionaries. Where this is not possible, we propose field values derived from controlled vocabularies. Since the database is of interest for the detection and reaction to disease outbreaks in wild animal populations, our model proposition does not encompass information pertaining to genetic sequencing that has no direct effect on the outcome of the laboratory result.

1.4.1. Capturing the host-pathogen relationship

Epidemiological data differ from wildlife distribution data in several ways, the most important difference being that each observation most commonly corresponds to the presence of two species – the host and the pathogen. This is not an uncommon scenario in biodiversity data – there are many instances where relationships exist between organisms observed at the same time, for example, the interaction between prey, predator, and primary producer. However, the relationship between host and pathogen that appears in epidemiological data is not in itself the point of interest when compared to, for example, the relationship between individuals in a pack, where the nature of the interaction can be both more complex and the focus of the research.

Determining the presence of a pathogen relies in its simplest form on physical examination of the host individual (for lesions, for example), and in its most complex on taking one or multiple tissue samples for analysis (excepting environmental DNA extraction, where the pathogen is disassociated from the host). A number of solutions currently exist in Darwin Core to handle such sampling situations. Firstly, two occurrences may be associated using the associatedOccurrences term (Table 3). In this instance, each occurrence of a pathogen could be associated with the occurrence of its host. However, the usage is somewhat limited, because the content of the resulting associatedOccurrences term becomes an

unwieldy, non-interoperable, concatenated string. Furthermore, the downside of this option is that these terms have an uncontrolled vocabulary, making them hard to standardise and query.

Table 3: example of the Occurrences extension employing the associatedOccurrences term

occurrenceID	scientificName	occurrenceStatus	associatedOccurrences
occ1	<i>Sus scrofa</i>	present	host of occurrence2
occ2	ASFv	present	virus collected from occurrence1

A second alternative is to use the resourceRelationship extension. The resourceRelationship is a widely adopted solution for describing relationships. This extension enables a relationship to be described between two or more resources that may be present in the event core, in one of the extensions, or even externally to the data set. The resourceRelationship extension is a powerful solution when the primary interest lies in the biological nature of the relationships. For example, this is particularly useful when we are interested in the relationship between individuals in a social group (e.g. parentOf, siblingOf) or when relating one data set with external data. For example, the relationship between a wild boar (*Sus scrofa*) and the African Swine Fever Virus (ASFv) could be described in the resourceRelationship table, whereby ASFv is the 'virusOf' *Sus scrofa* (relationship1) and *S. scrofa* is the 'hostOf' ASFv (relationship2, Table 4).

Table 4: example of the resourceRelationship table

resourceRelationshipID	resourceID	relatedResourceID	relationshipOfResource
relationship1	ASFv	<i>Sus scrofa</i>	virusOf
relationship2	<i>Sus scrofa</i>	ASFv	hostOf

However, the resourceRelationship table quickly becomes bulky and inelegant when the relationship between observations is not the focal interest of the data, or when the relationship between different entities is unknown. Similarly, the sequence and dependency of the observations is also lost, and becomes increasingly bulky when we include the host, the tissue sample and the pathogen.

Proposition

If we wish to preserve the essence of the host-pathogen relationship and the dependency of disease identification via a sample from the host, then neither the associatedOccurrences nor the resourceRelationship provides an elegant solution with sufficient clarity or sufficient room for providing details for both species' observations, respectively. However, a term exists for handling such sampling, the *preparations* term – "A list (concatenated and separated) of preparations and preservation methods for a specimen". Thus, a *preparation* (this could be in the form of a blood or a tissue sample) may be taken from the host species (Occurrence), and analyses conducted on the *preparation* for disease confirmation (Figure 2). To reinforce the difference and distinguish between the original host observation and the diagnosis of disease presence from a tissue sample, we suggest that the controlled vocabulary 'human observation' and 'material sample' be employed in the *basisOfRecord* term. Thus, for example, a blood sample preparation (*basisOfRecord:material sample*) is taken from a wild boar Occurrence (*basisOfRecord:human observation*) for confirmation of ASFv presence.

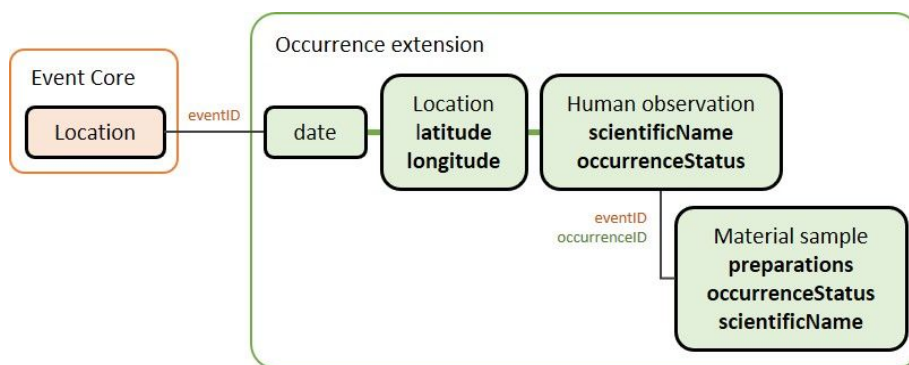


Figure 2: An Event takes place within a location (e.g. an administrative unit, a hunting ground, etc.). For each event, an observation of a host individual takes place at a precise time and location (the Occurrence). The Occurrence is the origin of a material sample (e.g. a blood sample), from which an identification of an animal disease is achieved.

For instance in the example given in Table 5a, capture events of wild boar for testing for the presence of African Swine Fever (ASFv, Table 5a) can be mapped to DwC thus: a location is surveyed for disease outbreak (Table 5b), wherein an occurrence of a host species happens on a given date (*basisOfRecord: human observation*), and a *preparation* extracted (*basisOfRecord: material sample*, Table 5c). Each preparation is analysed for disease diagnosis (*occurrenceStatus*). This is modelled by employing nestedness in the Occurrence extension, and use of the *parentOccurrenceID* term (Figure 2, Table 5c). It should be noted here that we have chosen to follow the OBIS guidelines of storing the quantity of host individuals in the extended MoF extension. This is to accommodate situations where the quantity of hosts observed cannot be identified as individuals, but instead through a weight or volume. In cases where the host individual is identifiable to the individual, this can be denoted thus: *measurementType: organismQuantity*, *measurementValue: 1*, *measurementUnit: individual* (Table 5d).

The notion of nestedness in Occurrences remains a general concept for a simple way to store dependent samples. To maintain coherence amongst the Darwin Core vocabulary, we suggest the use of the *parent* nomenclature, which has been adopted widely throughout Darwin Core and GBIF (for example, *eventID* and *parentEventID*, and *taxonID* and *parentTaxonID*, and under consideration for *measurementID* and *parentMeasurementID*).

Table 5a: raw data

species	nr observed	location	Date	blood sample taken	ASFv present
wild boar	1	Auriol	19/03/2020	yes	yes
wild boar	1	Auriol	22/03/2020	yes	yes
wild boar	1	Cabriès	30/01/2018	yes	no

Table 5b: event table

eventID	locality
event1	Auriol
event2	Cabriès

Table 5c: the occurrence table for a case of African Swine Fever Virus (ASFv) present in wild boar

occurrence ID	event ID	parent Occurrence ID	basisOf Record	latitude	longitude	scientific Name	organism Scope	preparations	occurrence Status	dateIdentified
event1/occ1	event1		human observation			Sus scrofa	multicellular organism		present	2020-03-19
event1/occ2	event1		human			Sus	multicellular		present	2020-03-22

event2/occ1	event2		observation		scrofa	organism			
			human observation		Sus scrofa	multicellular organism		present	2018-01-30
event1/occ1-1	event1	event1/occ1	material sample		ASFv	virus	blood sample	present	
event1/occ2-1	event1	event1/occ2	material sample		ASFv	virus	blood sample	present	
event2/occ1-1	event2	event2/occ1	material sample		ASFv	virus	blood sample	absent	

Table 5d: eMoF extension table where two material samples are extracted from the same host individual

measurementID	occurrenceID	eventID	measurement Type	measurement Value	measurement Unit
event1/occ1-1/meas1	event1/occ1-1	event1	organismQuantity	1	individual
event1/occ2-1/meas1	event1/occ2-1	event1	organismQuantity	1	individual
event2/occ1-1/meas1	event2/occ1-1	event2	organismQuantity	1	individual

1.4.2. Handling multiple laboratory analyses from one individual

In the case of some diseases it is standard procedure to perform multiple analyses and checks on the same host individual, providing a more robust diagnosis. The final diagnosis can often be the result of analyses originating from multiple locations on the host or multiple testing of the same sample, which can return either positive or negative to the presence of a pathogen.

One solution to handle this situation is to disregard the underlying analyses and report only the overall diagnosis in the occurrence extension. This would be a very simple solution, but would result in a lot of data loss on the detailed results from the specific analyses. Indeed, these details are a requirement for reporting disease outbreaks to the European Union (EU) for certain critical diseases. Finding a solution that meets the EU-standard data submission would be of great advantage for future liaison of wildlife disease data with host distribution data.

Proposition

Solutions that avoid data loss therefore, are preferable. An alternative solution could be to store the overall diagnosis in the Occurrence extension, and the detailed analysis results in the extended MeasurementOrFact (eMoF) extension, an extension capable of storing measurements or facts relating to any Darwin Core class. This would be modelled in a two-step process: firstly, each host occurrence would be associated with as many material sample *preparations* as are collected in a hierarchical parent-child Occurrence. Secondly, each material sample connected to a host occurrence could be associated to a measurement in the eMoF extension, where details of the results and the analyses would be stored as measurements. The advantage of nested occurrences is the one-to-many relationship between the parent Occurrence and the child-, or sub-occurrences. Thus, for each occurrence of the host, there may be as many pathogen identification procedures as there are material samples extracted.

Let us take the example of a common buzzard that tested positive for Avian Influenza (AI, Table 6a). The individual was observed at a precise location at a precise point in time and tested for the presence of AI by taking two swabs for lab analysis – one cloacal and one tracheal swab. Each swab was analysed using a PCR test to identify the presence of the virus, of which the cloacal swab returned negative, and the trachea swab positive.

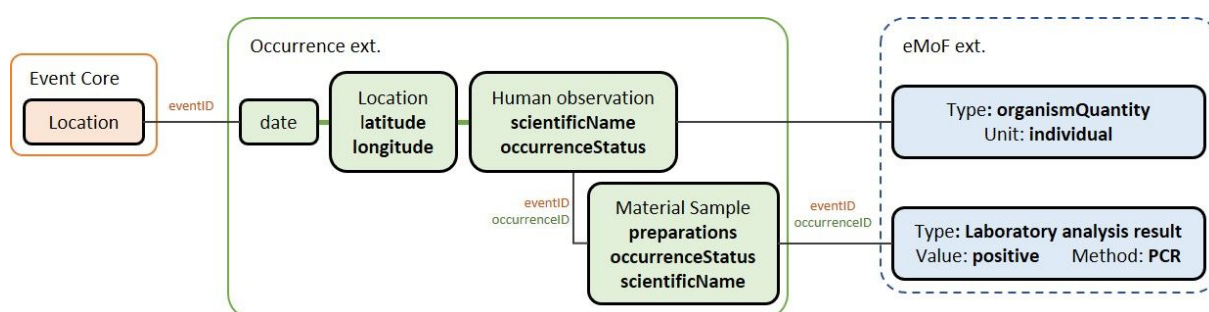


Figure 3: one host individual is sampled for disease identification, captured in a nested Occurrence structure. The relationship is one-to-many between the host occurrence and the tissue sample, accommodating multiple swabs to be taken from the same host.

Using the *preparations* term, information on the analysis for disease presence from different tissue samples collected from a host individual can be stored and traced. It should be noted here that the eMoF extension is capable of holding an infinite number of measurement types in the *measurementType* column, making it an extremely powerful extension, yet adding new terms should be carefully evaluated to avoid inflating and dispersing the number of terms. Thus a good practice would be at least to write terms in full letters, in singular, in miniscule and in english. Additionally, and of great interest to epidemiological data, a number of other terms beyond *measurementType* and *measurementValue* exist in the extension, namely *measurementUnit*, *measurementMethod*, *measurementAccuracy*, *measurementDeterminedBy*, *measurementDeterminedDate*, and *measurementRemarks* (Figure 4). For any given analysis value, a unit, method, operator, and date of determination can be supplied. This is particularly relevant for laboratory-based measurements, where the method used and the date of the analysis can be documented together with the test result (as illustrated in Figures 3 and 4, and Tables 6b-c).

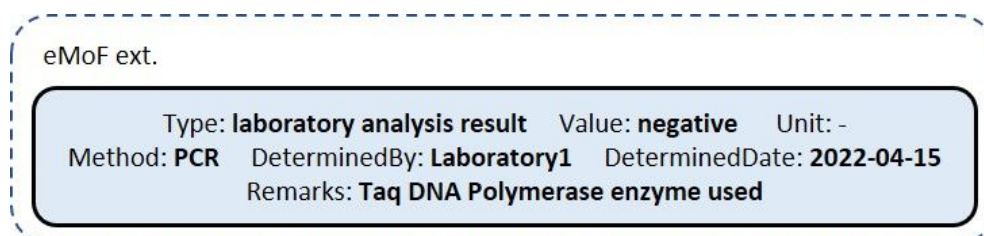


Figure 4: variables contained within the eMoF extension. 'measurement' has been dropped from the beginning of each term for illustration purposes.

Table 6a: raw data of a buzzard testing positive for highly pathogenic avian influenza (HPAI)

species	number of skin subsamples	subsample weight (g)	skin result PCR	trachea result PCR	analysed Laboratory1	polymerase enzyme
common buzzard	1	1	negative	positive	15/04/2022	taq

Table 6b: Occurrence extension where two material samples are extracted from the same host individual

occurrence ID	parent OccurrenceID	basisOf Record	preparations	scientific Name	organism Scope	occurrence Status
occ1		human observation		Buteo buteo	multicellular organism	present
occ1-1	occ1	material sample	trachea swab	AI	virus	present
occ1-2	occ1	material sample	skin	AI	virus	absent

Table 6c: eMoF extension table where two material samples are extracted from the same host individual

measurementID	occurrenceID	measurementType	measurementValue	measurementUnit	measurementMethod	measurementDeterminedBy
occ1/meas1	occ1	organismQuantity	1	individual		
occ1-1/meas1	occ1-1	laboratory analysis result	positive		PCR	Laboratory1
occ1-2/meas1	occ1-2	laboratory analysis result	negative		PCR	Laboratory1

Further complexity can be added when subsamples are extracted from preparations for analysis. This invokes another level in the hierarchical structure in the Occurrence extension, whereby a host individual is parent to a tissue sample *preparation*, the preparation in turn a parent to multiple child preparations. Continuing from the example above, a subsample of skin was taken from the original skin sample for disease verification, for which analyses were conducted only on the subsample (Figure 5, Tables 7a-b).

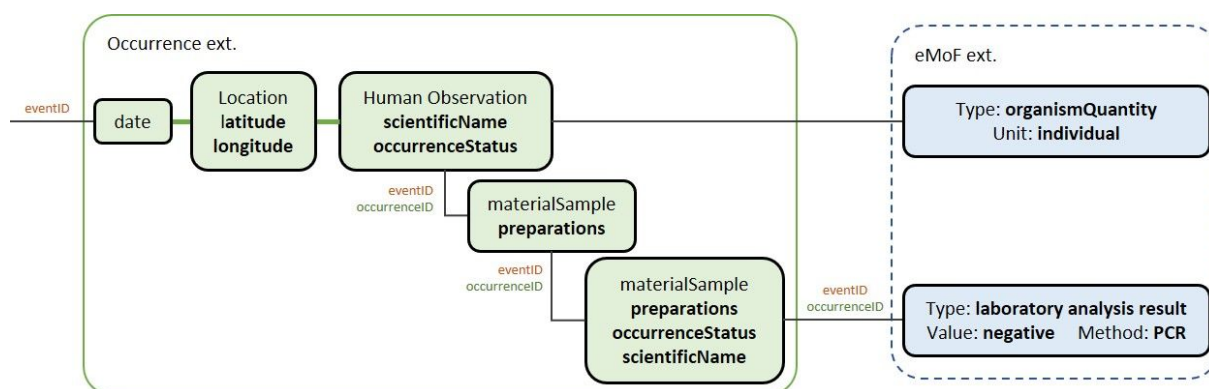


Figure 5: a subsample is extracted from a material sample collected from a host individual, resulting in a nested Occurrence structure

Table 7a: Occurrence table

occurrence ID	parentOccurrence ID	preparations	basisOfRecord	scientific Name	organism Scope	infraspecific Epithet	occurrence Status
occ1			human observation	Buteo buteo	multicellular organism		present
occ1-1	occ1	trachea swab	material sample	Avian Influenza virus	virus		present
occ1-2	occ1	skin	material sample	Avian Influenza virus	virus	H7N9	present
occ1-2-1	occ1-2	skin	material sample	Avian Influenza virus	virus	H7N9	absent

Table 7b: eMoF table

measurement ID	occurrence ID	measurement Type	measurement Value	measurementUnit	measurementMethod
occ1/meas1	occ1	organismQuantity	1	individual	
occ1-1/meas1	occ1	result	positive		PCR
occ1-2-1/meas1	occ1	result	negative		PCR

Here, subsamples can be taken from a *preparations* sample that was extracted from the host (Table 7a). Each subsample can be analysed separately using different laboratory techniques, for example, the information for which is stored in the eMoF table (table 7b). Each measurement associated with a *preparations* instance includes a *measurementValue*, a *measurementUnit*, and a *measurementMethod*.

We thus propose making use of the MeasurementOrFact extension and employing the *preparation* term in the Occurrence extension.

1.4.3. Handling multiple individuals in one laboratory analysis

When identifying the presence of a pathogen, samples taken from multiple host individuals can be pooled in the same laboratory analysis. Each observation of the host species does not therefore necessarily correspond to a unique observation of the pathogen.

In these instances, we cannot simply attribute one individual or one material sample to one test result. One solution could be to use controlled vocabulary in the extended MeasurementOrFact extension to relate the number of individuals pooled per analysis to each occurrence of the material sampled containing the pathogen. The problem here is that the information is relatively non-interoperable, and leaves ambiguity in the Occurrence extension as to how many of the observed individuals were analysed.

Proposition

Employing the Occurrence extension once again, however, we are able to specify how many individuals were sampled, independently of how many were observed. Take for instance a scenario where two individuals were observed (the Occurrence) and sampled, and both cloacal swab material samples (*preparations*) were pooled for analysis, but only one of the two individuals was sampled via tracheal swabs for analysis (Table 8a):

Table 8a: raw data denoting the number of individuals observed relative to the number of individuals pooled in a PCR analysis

sample cloaque	type	cloaca result PCR	trachea result PCR	nb individuals	nb individuals in same cloaca PCR	nb individuals in same trachea PCR
cloacal swab		negative	positive	2	2	1

Nevertheless, to reinforce the pooling effect, a measurementType 'number of individuals pooled' could be added to the eMoF extension (Table 8b-c).

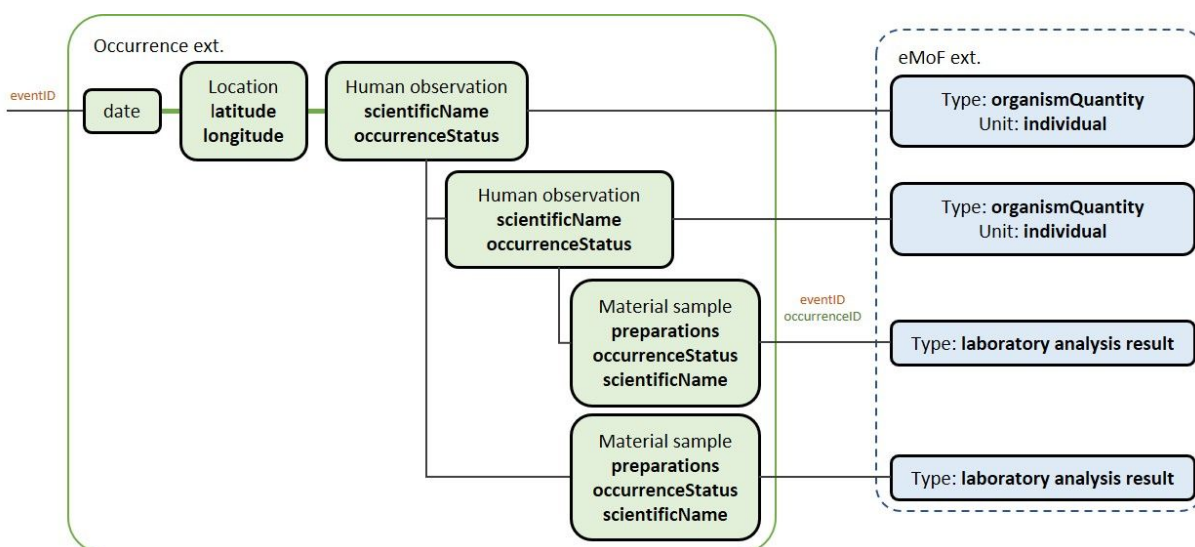


Figure 6: The flexibility provided by nesting material sample allows a subset of individuals to be sampled for further *preparations* (tracheal swabs).

Table 8b: nesting of host observations and material samples in the Occurrence extension

occurrence ID	parent OccurrenceID	basisOfRecord	preparations	scientific Name	organism Scope	occurrence Status
occ1		human observation		Buteo buteo	multicellular organism	present
occ1-1	occ1	human observation		Buteo buteo	multicellular organism	present
occ1-2	occ1	material sample	cloacal swab	Avian Influenza virus	virus	absent
occ1-1-1	occ1-1	material sample	trachea swab	Avian Influenza virus	virus	present

Table 8c: Extended measurementsOrFacts table denoting the laboratory analysis results and the number of individuals pooled in the sample

measurement ID	occurrence ID	measurementType	measurement Value	measurement Unit	measurement Method
occ1/meas1	occ1	organismQuantity	2	individual	
occ1-1/meas1	occ1-1	organismQuantity	1	individual	
occ1-2/meas1	occ1-2	laboratory analysis result	negative		PCR
occ1-2/meas2	occ1-2	number of individuals pooled	2		
occ1-1-1/meas1	occ1-1-1	laboratory analysis result	positive		PCR

1.4.4. Coping with virus strains and variants and their identification

Another difference between wildlife and epidemiological data is the nature of the species concerned. Pathogens, whether viral or bacterial, have a far higher mutation rate than those of bird and mammal species, for example. This fluid aspect of species taxonomy must be captured somehow, because often multiple variants of the same virus are tested for in the same sample. This is the case for Avian influenza,

where disease subtypes and variants are routinely tested for, and changes for each new epidemic. The Darwin Core model must, therefore, be capable of handling subspecies and variant nomenclature.

Proposition

The Taxon extension provides information on “a group of organisms considered by taxonomists to form a homogeneous unit”. Currently, Darwin Core accommodates scientific names under the term *scientificName*, where the complete scientific name can be recorded. Parts of the scientific name, such as subgenus (subgenus) or species epithet (specificEpithet), can be stored under additional terms. Pertinent for pathogen variants, however, is the term *infraspecificEpithet*, which handles “the name of the lowest or terminal infraspecific epithet of the scientificName, excluding any rank designation”. By allocating the pathogen variant to the *infraspecificEpithet*, we maintain general interoperability of the *scientificName*, whilst keeping the specificity of the variants separate. Each *preparations* can then be tested for (a). general disease presence, (b). a disease variant (Figure 7a), or (c). multiple disease variants (Figure 7b).

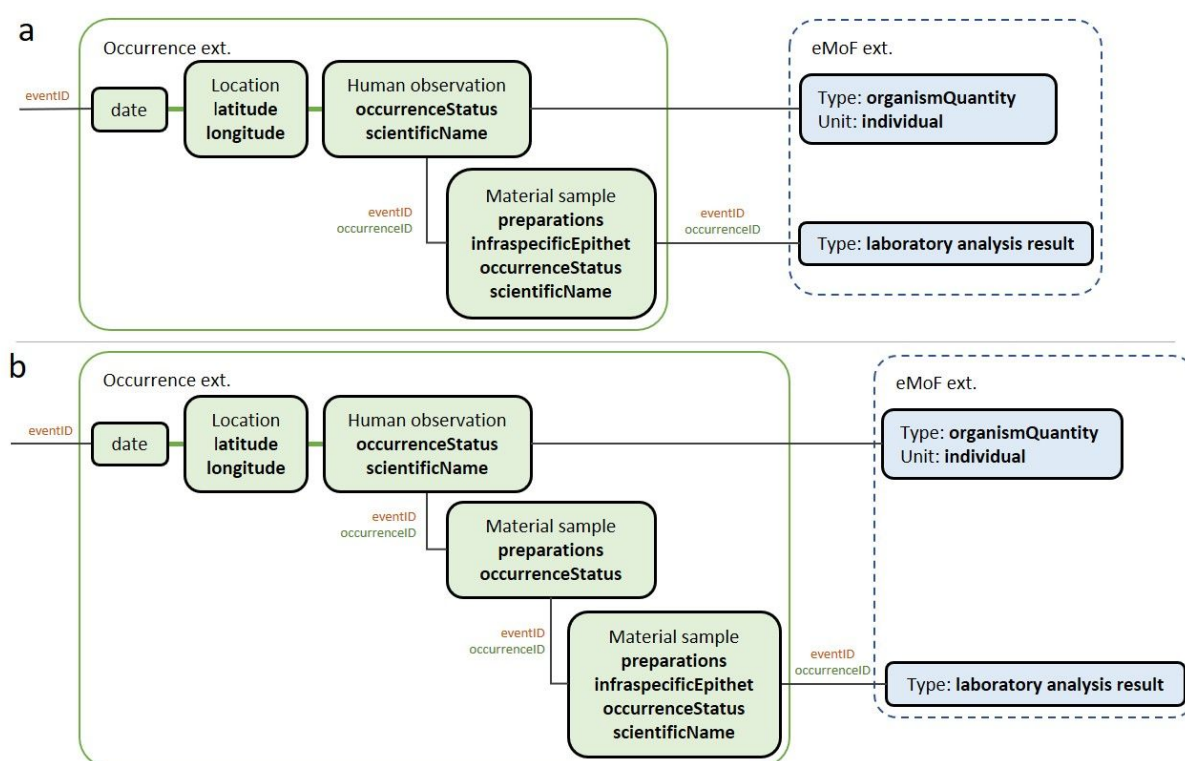


Figure 7: a. association of one material sample to one disease variant analysis; b. association of one material sample to multiple disease variant analysis

Using the example of the diagnosis of a mallard duck with the Avian Influenza virus (Table 9a), we can model the laboratory testing of two virus variants. Considering the first scenario depicted in Figure 7a, multiple material sample *preparations* are collected from the host individual (Table 9b), and each *preparations* is measured for disease diagnosis.

Table 9a: original data

species	sample type	result PCR	H5	H7	nb individuals
mallard duck	cloacal swab	positive	negative	positive	1

occurrence ID	parent OccurrenceID	scientific Name	basis Of Record	infraspecificEpithet	occurrenceStatus	preparations
occ1		Anas platyrhynchos	human observation			
occ1-1	occ1	Avian Influenza virus	material sample	H5N1	present	cloacal swab
occ1-2	occ1	Avian Influenza virus	material sample	H7N1	absent	cloacal swab

Table 9b: Occurrence table where two material samples are extracted from a single host individual for testing two viral strains, reported under infraspecificEpithet, see Figure 7a.

Secondly, as depicted in Figure 7b, a single material sample *preparations* is collected from the host individual, and analysed twice for the presence of two viral strains (Table 9c).

Table 9c: Occurrence table where one material sample is extracted from a single host individual for testing two viral strains, reported under infraspecificEpithet; see Figure 7b.

occurrence ID	parentOccurrenceID	scientific Name	basis Of Record	infraspecificEpithet	preparations	occurrence Status
occ1		Anas platyrhynchos	human observation			
occ1-1	occ1	Anas platyrhynchos	material sample		cloacal swab	
occ1-1-1	occ1-1	Avian Influenza virus	material sample	H5N1	DNA	present
occ1-1-2	occ1-1	Avian Influenza virus	material sample	H7N1	DNA	absent

1.4.5. Host data with details at the group level

Details relating to age, sex and reproductive condition of the host species can be recorded in the Occurrence extension. When more than one individual is observed and such details are available (or partially available), the Enetwild Wildlife Data Model¹ proposes using nested Occurrences. Let us imagine that three wild boars (*Sus scrofa*) were observed – one adult female, one adult male, and one with unknown age or sex. Following lab analyses of blood samples taken from the adult group members, the group was found to be positive for ASFv (Table 10a).

Table 10a: raw data denoting the observation of a group of wild boar that tested positive for African Swine Fever virus

species	nr observed	context	N female adult	N male adult	female adult result	male adult result
wild boar	3	ASFv	1	1	positive	negative

This structural data hierarchy can be captured by allowing the sub-occurrence of each adult individual tested (Table 10b) using parent-child notation in a one-to-many relationship. Using the flexibility created by the nested design, this level of complexity can be easily modelled. We propose a hierarchical occurrence structure, and the employment of the parentOccurrenceID term (proposed by Enetwild Consortium et al.¹). Such a structure would contain a parent observation, in this case of a host species, corresponding to an occurrence ID, and a child observation of the sub-group via a parentOccurrenceID. We rely on identical syntax to that of the parent-child relationship between the eventID and the parentEventID. In the example below, occurrenceID:occ1 of 3 wild boar individuals is comprised of 1 adult female (occurrenceID:occ1-1) and 1 adult male (occurrenceID:occ1-2). Both of these occurrences are child occurrences of occurrenceID:occ1, which is illustrated in the term parentOccurrenceID:occ1 (Table 10b).

Table 10b: proposed Occurrence extension where Occurrences are nested

occurrence ID	parent Occurrence ID	basis Of Record	preparations	scientific Name	infra specific Epithet	individual Count	occurrence Status	life Stage	sex
occ1		human observation		Sus scrofa		3	present		
occ1-1	occ1	human observation		Sus scrofa		1	present	adult	female
occ1-2	occ1	human observation		Sus scrofa		1	present	adult	male
occ1-1-1	occ1-1	material sample	blood sample	ASFv	subtype 1		present		
occ1-2-1	occ1-2	material sample	blood sample	ASFv	subtype 2		absent		

1.4.6. Avoiding duplication of dates

The dates at which events occur differ for the observation of the host species, and at which the presence of the virus was identified. For example, a pathogen might be identified at a lab four days after the host was observed and a sample taken and refrigerated or frozen. This puts into question the structure of the event. Does the event begin on the date on which the host was observed, and end on the date the virus was declared present or absent? Or rather do these dates correspond to the Occurrence of the host?

One solution could be to record the date of the event within the Event core, and store all additional dates in the measurementOrFact extension. Although simple, this solution makes data extraction by date more complex and less intuitive. Further, dates pertaining to the Event core relate to pre-meditated instances, such as a field season, not to spontaneous species observations such as the discovery of a wild boar carcass.

Another solution could be to make use of the measurementDeterminedDate term in the measurementOrFact extension – the date on which the measurementOrFact was made (http://rs.gbif.org/extension/obis/extended_measurement_or_fact.xml). This solution appears to be highly useful for associating the dates of analysis to the analysis outcome.

A further option would be to make use of the Identification class. Information on when the virus was identified, and by whom, can be neatly stored here, via the dateIdentified term. However, this bloats somewhat the number of terms needed in the Occurrence extension. Further, it would be impossible to store dates in cases where multiple analysis measurements are conducted on different dates but originating from the same material sample *preparations* item.

Proposition

We propose a multi-angle approach to handling the existence of multiple dates. Firstly, dates that refer to the timing of events (i.e. the observation process that leads to an occurrence observation) should be stored as eventDates in the Event core. This corresponds to the duration of a pre-meditated event, or field or hunting season, such as the duration of a transect. Dates corresponding to the exact observation of the host species can be stored in the Occurrence extension, as it corresponds to the exact timing of the observation, such as a specific time along a transect (conforming in this case to the universal format). If a sample is later taken from the host individual, the date can be informed in the occurrence table, corresponding to the material sample record. Dates that refer to later reception of samples by

the lab can be stored as measurements in the eMoF extension using a specific measurementType vocabulary (e.g. laboratory reception date), as it corresponds to very technical/administrative information. Finally, the date of disease diagnosis can be reported as *measurementDeterminedDate* in the measurementOrFact extension, in addition to more precise determination for each sample.

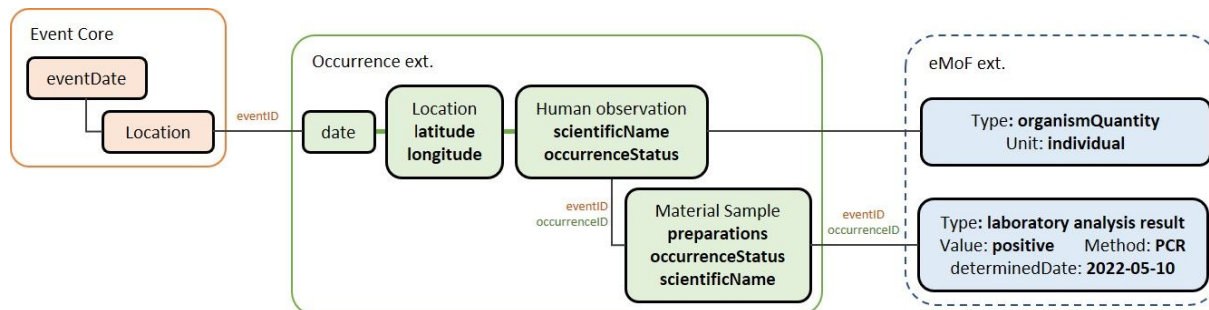


Figure 8: dates can be stored within the Event core, the Occurrence extension, or the extended MoF extension, according to the instance to which they are associated.

1.4.7. Statistical data

Where data are extracted from analyses or publications, the data are generally processed to a certain extent, notably through statistical analyses. Enetwild Consortium et al.¹ have already demonstrated that storing statistical data for wildlife distribution and abundance data is feasible in Darwin Core by employing a nested structure in the extended MeasurementsOrFacts table.

Proposition

Taking the example of reporting rabies outbreaks in bat colonies⁷ (Table 11a), the percentage of diseased individuals is reported together with an upper and lower confidence interval at 95%, which statistically reflect the 2.5% and 97.5% quantiles. These confidence intervals are dependent on the parent measurement (Table 11b), incurring a nested data structure.

Table 11a: raw data for rabies outbreaks in bats in 2009

date	% with EBLV-1 antibodies	95% CI
7 July	45.5% (15/33)	28.5–63.4
10 July	59.5% (22/37)	42.1–75.2

Table 11b: event core

eventID	eventDate
event1	2009-07-07
event2	2009-07-10

Table 11c: occurrence extension

occurrenceID	eventID	parent OccurrenceID	scientificName	occurrence Status
event1/occ1	event1		Eptesicus serotinus	present
event2/occ1	event2		Eptesicus serotinus	present
event1/occ1-1	event1	event1/occ1	European bat 1 lyssavirus	present
event2/occ1-1	event2	event2/occ1	European bat 1 lyssavirus	present

Table 11d: eMoF extension table for rabies outbreaks in bats

measurement ID	event ID	occurrence ID	parent Measurement ID	measurement Type	measurement Value	measurement Unit	measurement Method
event1/occ1/meas1	event1	event1/occ1		prevalence	45	percentage	antibodies
event1/occ1/meas2	event1	event1/occ1		organismQuantity	33	individual	
event2/occ1/meas1	event2	event2/occ1		prevalence	59.5	percentage	antibodies
event2/occ1/meas2	event2	event2/occ1		organismQuantity	37	individual	
event1/occ1/meas1-1	event1	event1/occ1	event1/occ1/meas1	x_0.025	28.5	percentage	
event1/occ1/meas1-2	event1	event1/occ1	event1/occ1/meas1	x_0.975	63.4	percentage	
event1/occ1/meas1-3	event1	event1/occ1	event1/occ1/meas1	sample size	33	individual	
event2/occ1/meas1-1	event2	event2/occ1	event2/occ1/meas1	x_0.025	42.1	percentage	
event2/occ1/meas1-2	event2	event2/occ1	event2/occ1/meas1	x_0.975	75.2	percentage	
event2/occ1/meas1-3	event2	event2/occ1	event2/occ1/meas1	sample size	37	individual	

1.4.8. Methodological and Sampling Information

Information on the diagnostics methodology and sampling strategy can be stored either in the data themselves, or in the metadata. Which of these two options is chosen for methodological and sampling information depends on the scale to which the information corresponds. Information associated with the entire dataset, such as the EU programme or European registration number or sampling strategy, can be reported in the metadata. Methodological information relative to part of the data set, such as analysis methods, can be reported in the data. This is not fixed, however, and sampling and analysis information can be stored in metadata or in the data itself depending on what part of the dataset it relates to. It is good practice to design a dataset as coherently as possible and derived from easily defined methodology in metadata. However, for interoperability reasons, there is an advantage to keeping some of the broader-scale methodological and sampling information in the data as opposed to the metadata. As a compromise, we propose to maintain a metadata structure similar to the existing metadata structure of the Enetwild Wildlife Database (Figure 9), whilst including the collection ID, the sampling protocol information in the data to facilitate the interoperability and data selection.

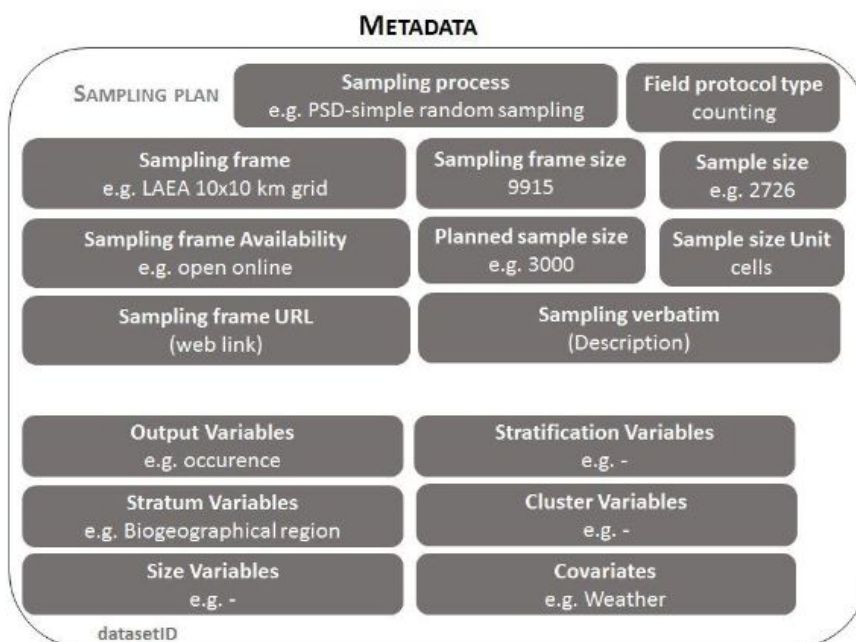


Figure 9: extracted from Enetwild Consortium et al., 2020.

2. Formatting propositions for epidemiological data

Given the problems and their potential solutions above, we suggest the following proposition for implementing the Darwin Core standard with epidemiological data. The suggestions benefit from using existing Darwin Core terms and terms previously proposed in the Enetwild Wildlife Data Model, without the need for further developments to the standard, thus following the original Darwin Core philosophy of simplicity.

To accommodate the novel complexities found in epidemiological data outlined above, we propose the following primary adjustments:

- Introduction of the ***parentMeasurementID*** term to handle nestedness in measurements
 - usage example: statistical or methodological details
- Introduction of the ***parentOccurrenceID*** term to incorporate nested occurrences prior to sampling, and to accommodate subsampling situations
 - usage example: partial information on the group of host individuals
 - usage example: Pathogen-host relationship, complemented by ***preparations*** to define the material in which the pathogen was identified
- Use of ***infraspecificEpithet*** to distinguish between pathogen subtypes
 - usage example: multiple pathogen strains tested for in one host individual
- Use of ***measurementMethod***, to provide details on laboratory analyses
- Use of ***measurementDeterminedBy***, to provide details on laboratory details
- Use of ***measurementDeterminedDate*** to define the date on which infection was determined

The proposition of nestedness within the eMoF extension and the Occurrence extension is a promising development for handling complex epidemiological data. Such hierarchical structuring is already adopted in some of the Global Biodiversity Facility (GBIF) data model extensions, demonstrating that this solution is powerful and safe for addressing dependencies in extensions.

A collecting Event happens at a Location during a period of time (included in the Event), and may follow a specified samplingProtocol with respect to the target Organism, (an Occurrence). The whole Organism or parts from it (*preparations*) may be gathered at a precise Location and at a precise point in time. Multiple Measurements may be associated with the *preparations*, the Occurrence, and the Event. A genetic sequence or Taxonomic identification is derived from the material sample *preparations*, which in turn is derived from an Organism or part thereof (a *preparations*) that was derived from a material sample *preparations*.

The Epidemiology Data Model thus relies on the following tables:

Event: eventDate, eventID, eventRemarks, eventType, parentEventID

Location (used in both the Event and Occurrence tables): coordinatePrecision, coordinateUncertaintyInMeters, country, countryCode, decimalLatitude, decimalLongitude, geodeticDatum, locality, locationID, locationAccordingTo, verbatimCoordinates, footprintWKT

Occurrence: basisOfRecord, individualID, infraspecificEpithet, lifeStage, occurrenceID, occurrenceStatus, organismQuantity, organismQuantityType, organismScope, parentOccurrenceID (**new**), preparations, recordedBy, recordType, reproductiveCondition, scientificName, sex

Extended MeasurementOrFact: measurementDeterminedDate, measurementID, measurementMethod, measurementType, measurementValue, measurementUnit, occurrenceID, parentMeasurementID (**new**)

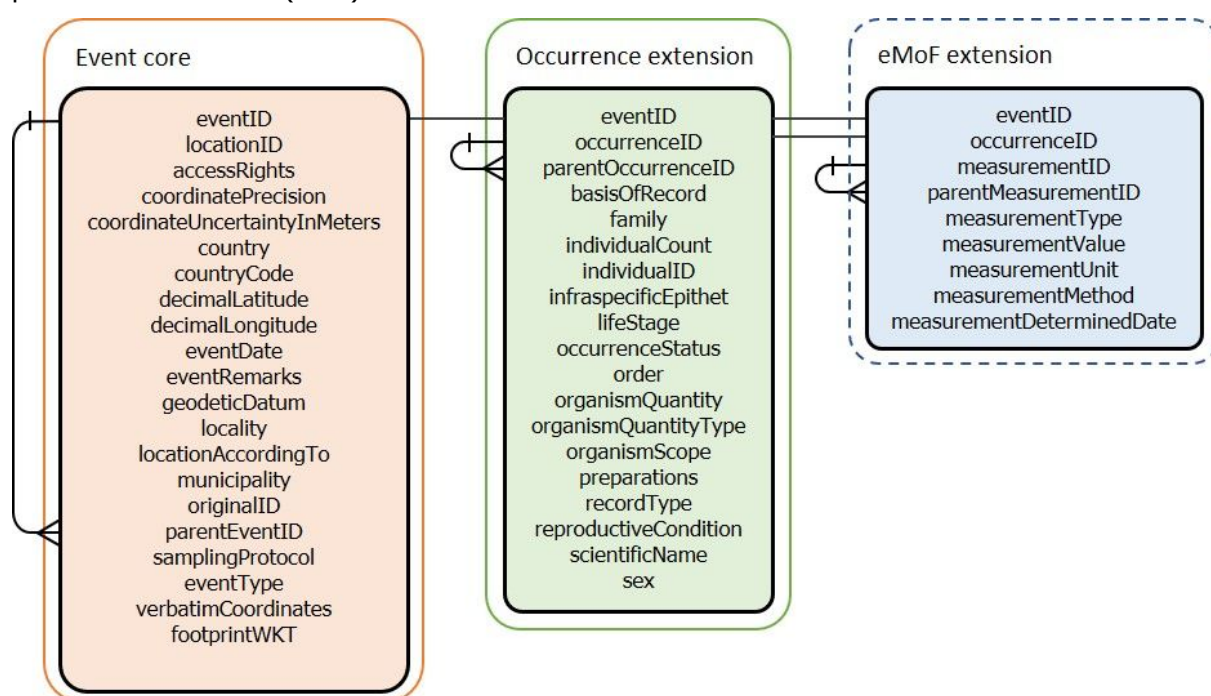


Figure 10: conceptual model for the Enetwild Wildlife Disease data model, consisting of (1) the Event Core containing information about the general sampling design, timing and location, (2) the Occurrence extension, where information on the host observation, the tissue sample, and identification of the pathogen, and (3) the eMoF extension, containing information pertaining to *either* the Event or the Occurrence, or both.

3. Case studies

Below we apply the proposed model to a number of case studies, ranging from national to European-scale surveillance programmes for wildlife disease to see whether the proposed model can accommodate the various current models related to this field or if it encounters limitations.

3.1. CSF/ASF EU database

A working group on Classical Swine Fever (CSF) in wild boar of the Directorate General Health and Consumer Protection ([DG SANCO](#)) of the European Commission (EC) has established a database on the epidemiology of CSF and African Swine Fever (ASF) for wild boar populations in all European member states and neighbouring countries (http://public.csf-wildboar.eu/Docs/EURL-CSF_ASF-WB-DB_Saxony_public.pdf and <http://public.csf-wildboar.eu/Default.aspx>).

The flat table database consists of 1 record for each wild boar shot or found dead. Each record contains a unique identification of wild boar, municipality, date of shooting/finding and age, in addition to details of the laboratory results (virology, serology), and an official confirmation by a competent authority (Table 12a).

Table 12a: column names for the CSF/ASF database with associate Darwin Core terminology

Field name	Description	Darwin Core table	Darwin Core term
AgeCode	Age class	occurrence	lifeStage
AreaCode	Level 3 key (usually municipality level)	event	locationID
CarcassCode	Type of carcass	eMoF	measurementType
Confirmed	Confirmation of the record by authorities, only confirmed Records can be analysed by other Participants	eMoF	measurementType
CountryCode	Country of sample origin	event	countryCode
HDate	Date of shooting/finding of each wild boar OR Date of Laboratory registration	occurrence	date
ID	Unique identification of the single wild boar (e.g. 2002/000372355)	occurrence	individualID
IsolatId	Highest homology found with known CSF isolate/strain	eMoF	measurementType
LabCode	Unique identification of the laboratory	eMoF	measurementType
MolHomology	Homology in % of most similar CSF isolate or strain, if known	eMoF	measurementType
RestrictionCode	Type of restriction area	eMoF	measurementType
ResultCode	Official judgement for CSF	occurrence	occurrenceStatus
SequenceMethodCode	Analyzed sequence of the CSF virus genome	eMoF	measurementType
SequenceNumber_ASF	Identification of the ASF molecular sequence	eMoF	measurementType
SerologyCode	Result of the CSF serological investigation	eMoF	measurementType

Wildlife disease data model standards

SertestCode	CSF serological test	eMoF	measurementMethod
SexCode	Sex	occurrence	sex
Vacc	CSF vaccination practised?	eMoF	measurementType
VirologyCode	Result of CSF virological investigation	eMoF	measurementType
VirtestCode	CSF virological test	eMoF	measurementType
VirtypeCode	CSF virus type	occurrence	infraspecificEpithet
XValue	X-coordinate of the position (decimal degree in WGS1984, e.g. GPS)	event	decimalLatitude
YValue	Y-coordinate of the position (decimal degree in WGS1984, e.g. GPS)	event	decimalLongitude

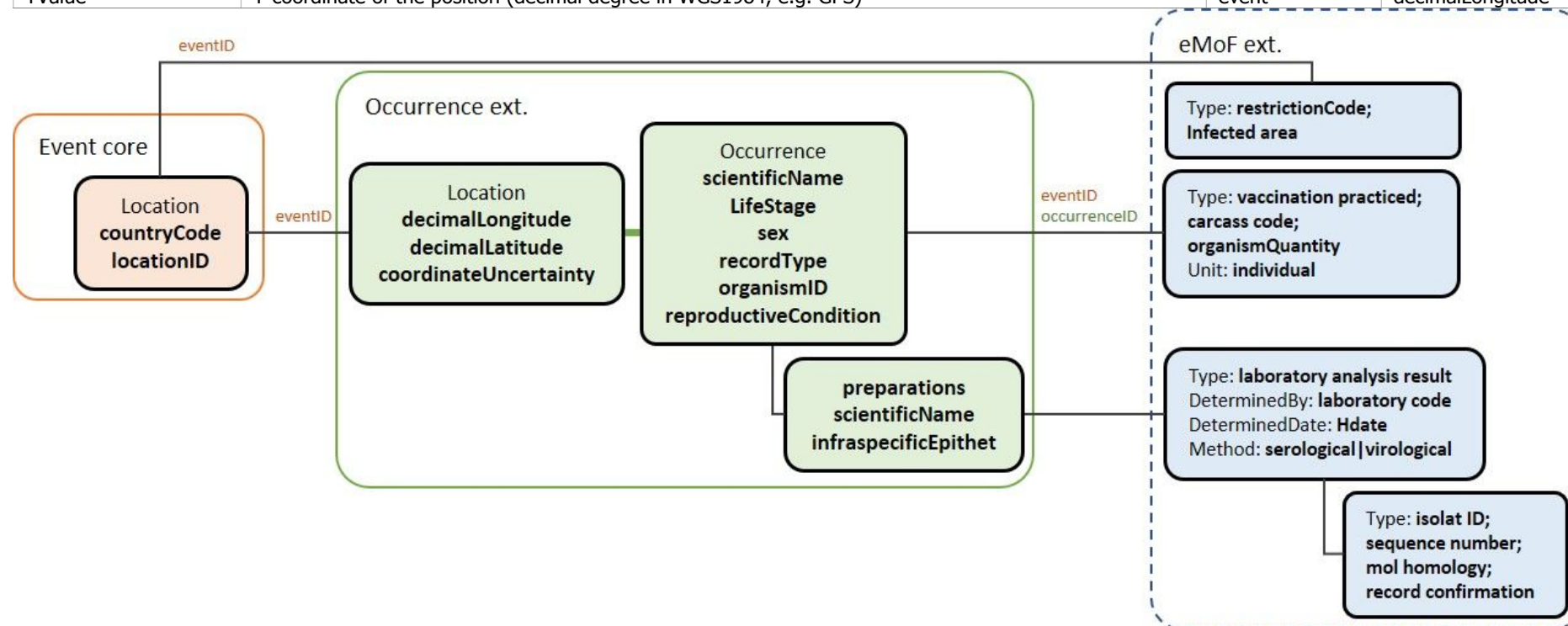


Figure 11: Conceptual model for the CSF/ASF EU database viewed under the DwC structure

The data presented above can be modelled in Darwin Core, with the geographical information referring to the event (Event) and their details as measurements in the extended MeasurementOrFact table, information about the carcass in the Occurrence extension, with its ID and vaccination status as measurements in the MeasurementOrFact table, and information on the material sampled and the virus presence (or absence) in a nested Occurrence, with result details as measurements in the MeasurementOrFact table.

Table 12b: the event Core for the CSF/ASF database

event ID	parentEventID	collectionID	accessRights	rightsHolder	habitat	locality	country	country Code	location ID	location AccordingTo	municipality	coordinateUncertaintyIn Meters	event Date
event1						areaCode		countryCode					

Table 12c: the occurrence extension for the CSF/ASF database

occurrenceID	eventID	parentOccurrenceID	basisOf Record	organism ID	kingdom	preparations	scientific Name	organismScope	infraspecificEpithet	occurrenceStatus	dateIdentified	life Stage	sex	reproductive Condition	record Type
event1/occ1	event1		human observation					multicellular organism		resultCode	Hdate	ageCode			
event1/occ1-1	event1	event1/occ1	material sample			sample		virus	virusType						

Table 12d: the extended MoF extension for the CSF/ASF database

measurementID	event ID	occurrence ID	parent MeasurementID	measurement Type	measurement Value	measurement Unit	measurement Method	measurement Determined Date	measurement Determined By
event1/meas1	event1			restrictionCode					
event1/occ1/meas1	event1	event1/occ1		carcass code					
event1/occ1/meas2	event1	event1/occ1		vaccination practiced					
event1/occ1/meas3	event1	event1/occ1		year					
event1/occ1/meas4	event1	event1/occ1		month					
event1/occ1/meas5	event1	event1/occ1		day					

Wildlife disease data model standards

event1/occ1/meas6	event1	event1/occ1		record confirmation					
event1/occ1/meas7	event1	event1/occ1		organismQuantity		individual			
event1/occ1-1/meas1	event1	event1/occ1-1		laboratory analysis result			virology serology	Hdate	labCode
event1/occ1-1/meas1-1	event1	event1/occ1-1	event1/occ1-1/meas1	sequence number					
event1/occ1-1/meas1-2	event1	event1/occ1-1	event1/occ1-1/meas1	mol homology					
event1/occ1-1/meas1-3	event1	event1/occ1-1	event1/occ1-1/meas1	isolat ID					

3.2. Tuberculosis in UK badgers

A long-term study of a wild badger population in an area of about 7sq km in Southwest England has been conducted annually since the 1970s. The research involves regular trapping and release of badgers, and testing for tuberculosis using the DPP, IGRA blood tests and culture. The database contains a variety of variables corresponding generally to a. the timing and location of the sampling event, b. the timing, location and condition of the observed individual, c. the type of sample(s) collected for examination, and d. analyses conducted on the sampled material (Table 13a).

The flat table database consists of 1 record for each captured badger. Each record contains a unique badger identification, badger sett name, date of (first) capture and age, the condition of the individual, in addition to details of the laboratory results (Table 13a).

Table 13a: an overview of the columns stored in the badger tuberculosis dataset

Field Name	Explanation	darwinCore table	darwinCore term
Abscesses	Description of abscess	eMoF	measurementType
age	Age of wound e.g. fresh or scar	eMoF	measurementType
age_fc	Age of the badger when first captured. Split into categories (cub, yearling, adult)	occurrence	lifeStage
amount	Number of wounds	eMoF	measurementType
area	Wounded area of the badger e.g. Limb, neck	eMoF	measurementType
badgerId	The ID given to the badger	eMoF	measurementType
Blood	Has a blood sample been taken?	eMoF	measurementType
body_length	Length of body	eMoF	measurementType
captureDate	The date that the badger was captured	entity	Date
CaptureYear	Year of capture	event	eventDate
Condition	The condition of the badger when captured	eMoF	measurementType
Date	The date when the animal was cultured	eMoF	Date
Date	The date when ELISA test was carried out	eMoF	measurementDeterminedDate
Experiment_Date	Date of ELISA experiment	eMoF	measurementDeterminedDate
hair_sample	Was a hair sample taken?	eMoF	measurementType
Lesions	Are there visible lesions? Y/N/NA	eMoF	measurementType
Mean_OD	Mean od results in full	eMoF	measurementType
neck_diameter	Neck diameter	eMoF	measurementType
OpticalDensity1	0 – 1,000,000		measurementType
OpticalDensity2	0 – 1,000,000	eMoF	measurementType
os_ref	Map reference of where the badger was found or captured	event	decimalLatitude decimalLongitude
PlateNo	Sample plate ID	eMoF	measurementType
pm_cause	Cause of death	eMoF	measurementType
Pmstat	Status of post-mortem	eMoF	measurementType
PPD-A_1	PPD Avian 1 st reading	eMoF	measurementType
PPD-A_2	PPD Avian 2 nd reading	eMoF	measurementType
PPD-B_1	PPD Bovine 1 st reading	eMoF	measurementType
PPD-B_2	PPD Bovine 2 nd reading	eMoF	measurementType
Ref	ELISA reference	eMoF	measurementType
Reproc	Reproductive condition, females only	occurrence	reproductiveCondition
Result	Result code – negative/positive/inconclusive/borderline/no sample	eMoF	measurementType
Retest	Was a retest required ?	eMoF	measurementType
Sample	Type of sample	occurrence	preparations
SampleType	Serum, Whole Blood, Whole Blood (Field)	occurrence	preparations
Sett	Name of sett	eMoF	measurementType
Sex	Sex of the badger. Male/Female	entity	sex
size	Size of wound	eMoF	measurementType
Socg	Social group	eMoF	measurementType
Spoligo	Spoligotype	eMoF	measurementType
StatPakId	P-Positive, N-Negative, C-No Clotted sample	eMoF	measurementType
Tattoo	The tattoo id that the animal was registered with	occurrence	individualID

Temp	Body temperature	eMoF	measurementType
Test_Failed	The reason why the test failed	eMoF	measurementType
Test_Info	Information on the outcome of the test	eMoF	measurementRemarks
Testes	Description of state of testes, males only	entity	reproductiveCondition
TestType	Double or Single Band	eMoF	measurementMethod
TimeBetweenCapture	Max time in months between trappings. If only trapped once then 0.	eMoF	measurementType
Toothwear	Degree of toothwear	eMoF	measurementType
VisualLine1	[P]ositive or [N]egative	eMoF	measurementType
VisualLine2	[P]ositive or [N]egative	eMoF	measurementType
Weight	Weight of the badger when captured.	eMoF	measurementType
WhiskerTaken	Whisker sample taken	eMoF	measurementType
year_fc	Year when the badger was first captured	eMoF	measurementType

The data presented above can be modelled in Darwin Core, with the geographical information referring to the event (Event, Table 13b) and their details as measurements in the extended MeasurementOrFact table, information about the trapped badger in the Occurrence extension (Table 13c), with its ID as measurements in the MeasurementOrFact table, and information on the material sampled and the pathogen presence (or absence) in a nested Occurrence, with result details as measurements in the MeasurementOrFact table (Table 13d).

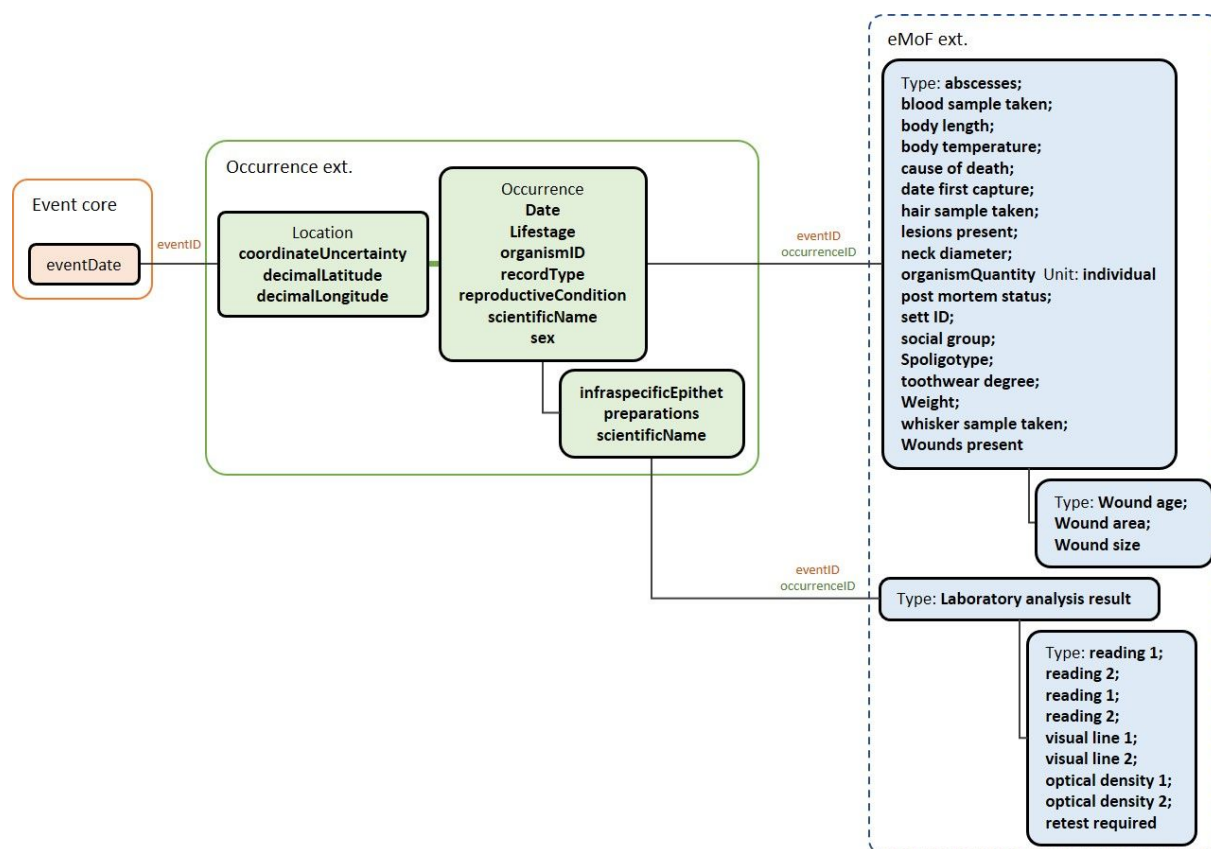


Figure 12: conceptual figure for the long-term surveillance study of tuberculosis in badgers

Table 13b: conceptual model – the Event Core

event ID	parent EventID	collection ID	access Rights	rights Holder	habitat	locality	country	country Code	location ID	location According To	municipality	event Date
event1							UK	UK	SW			1970

Wildlife disease data model standards

Table 13c: occurrence table

event ID	occurrence ID	parent Occurrence ID	organism ID	basisOf Record	preparations	scientificName	organism Scope	infraspecific Epithet	occurrence Status	life Stage	sex	reproductive Condition	coordinate Uncertainty	decimalLatitude	decimalLongitude	record Type
event1	event1/occ1		ind9872	human observation		Meles meles	multicellular organism			adult	female			lat	lon	found dead
event1	event1/occ1-1	event1/occ1		material sample	whole blood	Mycobacterium bovis	virus		present							

Table 13d: eMoF extension. Note that the term 'measurement' has been removed from the headings in order to increase readability. Thus, 'Type' becomes 'measurementType', etc.

measurement ID	event ID	occurrenceID	parent MeasurementID	Type	Value	Unit	Method	Determined Date	Determined By	Remarks
event1/occ1/meas1	event1	occurrence1		abscesses	text					
event1/occ1/meas2	event1	occurrence1		lesions present	yes no					
event1/occ1/meas3	event1	occurrence1		spoligotype						
event1/occ1/meas4	event1	occurrence1		wounds present	yes no					
event1/occ1/meas5	event1	occurrence1		body length	80	cm				
event1/occ1/meas6	event1	occurrence1		weight	40	kg				
event1/occ1/meas7	event1	occurrence1		date first capture	06/03/2003	yyyy-mm-dd				
event1/occ1/meas8	event1	occurrence1		body temperature						
event1/occ1/meas9	event1	occurrence1		blood sample taken						
event1/occ1/meas10	event1	occurrence1		toothwear degree						
event1/occ1/meas11	event1	occurrence1		whisker sample taken	yes no					
event1/occ1/meas12	event1	occurrence1		hair sample taken	yes no					
event1/occ1/meas13	event1	occurrence1		social group						
event1/occ1/meas14	event1	occurrence1		sett ID						
event1/occ1/meas15	event1	occurrence1		post mortem status						
event1/occ1/meas16	event1	occurrence1		cause of death						
event1/occ1/meas17	event1	occurrence1		neck diameter		cm				
event1/occ1/meas18	event1	occurrence1		organismQuantity		individual				
event1/occ1-1/meas1	event1	occurrence1-1		result	positive negative		IIFN Gamma ELISA	yyyy-mm-dd		test info
event1/occ1/meas4-1	event1	occurrence1	event1/occ1/meas4	age	scar					
event1/occ1/meas4-2	event1	occurrence1	event1/occ1/meas4	number						
event1/occ1/meas4-3	event1	occurrence1	event1/occ1/meas4	area	limb					
event1/occ1/meas4-4	event1	occurrence1	event1/occ1/meas4	size	26	mm				
event1/occ1-1/meas1-1	event1	occurrence1-1	event1/occ1-1/meas1	reading 1			PPD - Avian	yyyy-mm-dd		
event1/occ1-1/meas1-2	event1	occurrence1-1	event1/occ1-1/meas1	reading 2			PPD - Avian	yyyy-mm-dd		
event1/occ1-1/meas1-3	event1	occurrence1-1	event1/occ1-1/meas1	reading 1			PPD - Bovine	yyyy-mm-dd		

Wildlife disease data model standards

event1/occ1-1/meas1-4	event1	occurrence1-1	event1/occ1-1/meas1	reading 2			PPD - Bovine	yyyy-mm-dd		
event1/occ1-1/meas1-5	event1	occurrence1-1	event1/occ1-1/meas1	visual line 1	positive negative		DPP - double band	yyyy-mm-dd		test info
event1/occ1-1/meas1-6	event1	occurrence1-1	event1/occ1-1/meas1	visual line 2	positive negative		DPP - single band	yyyy-mm-dd		
event1/occ1-1/meas1-7	event1	occurrence1-1	event1/occ1-1/meas1	optical density 1	0 – 1,000,000					
event1/occ1-1/meas1-8	event1	occurrence1-1	event1/occ1-1/meas1	optical density 2	0 – 1,000,000					
event1/occ1-1/meas1-9	event1	occurrence1-1	event1/occ1-1/meas1	retest required	yes no					

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.3. SIGMA database

The SIGMA laboratory database stores information on laboratory results of disease surveillance in wildlife and domestic animals. It has so far been developed and used for African Swine Fever virus, and is currently undergoing adjustment for Avian Influenza. The European Food Safety Authority (EFSA) uses the SIGMA laboratory data model to collect information in a harmonised way from EU Member States and neighbouring countries on their national disease surveillance programs for the development of annual epidemiological reports.

The database consists of information on the sampling programme, the sampling event, the sample taken, the analytical method employed, and the analysis result for a given host individual. An occurrence of one or more instances of the host species provides one or more *preparations* for laboratory testing. Tests can be conducted for one or many virus variants on a singular or pooled sample, or on portions of a single sample. Further information on the quantitative laboratory results are recorded.

Table 14a: the SIGMA data model

SIGMA terms	Hierarchical arrangement	DarwinCore terms	controlled vocabulary
sampEventID	Event	eventID	
	Event	parentEventID	
progID	Event	collectionID	
progType	Event	accessRights	
progLegalRef	Event	rightsHolder	
sampPoint	Event	habitat	
progInfo.targetGroup	Event	recordType	
sampUnitIds.sampHuntingGround	Event	locality	
sampCountry	Event	country	
sampArea	Event	locationID	
sampMunicipality	Event	municipality	
sampInfo.coordPrecision	Occurrence	coordinateUncertaintyInMeters	
sampInfo.latitude	Occurrence	decimalLatitude	
sampInfo.longitude	Occurrence	decimalLongitude	
	Occurrence	eventID	
sampUnitType	Occurrence	kingdom	
sampUnitIds.animalId	Occurrence	organismID (individual identification number)	
sampId	Occurrence	occurrenceID	
sampMatType	Occurrence	ggbn:materialSampleType	
sampMatCode.part	Occurrence	preparations	
sampMatCode.source	Occurrence	scientificName	
	Occurrence	organismScope	
paramCode	Occurrence	infraspecificEpithet	
sampSize	Occurrence	organismQuantity	
sampSizeUnit	Occurrence	organismQuantityType	
resQualValue	Occurrence	occurrenceStatus	
sampY	Occurrence	year	
sampM	Occurrence	month	
sampD	Occurrence	day	
sampMatCode.animage	Occurrence	lifeStage	
sampMatCode.gender	Occurrence	sex	
	Occurrence	identifiedBy	
anMethRefId	Occurrence	samplingProtocol	
progInfo.com	eMoF	measurementType	description of goal
evalInfo.nationalCaseId	eMoF	measurementType	ADNS ADIS ID number
sampMatCode.prod	eMoF	measurementType	animal category
sampMatText	eMoF	measurementType	carcass decomposition
anMethCode	eMoF	measurementType	anMethCode
anMethInfo.methSensitivity	eMoF	measurementType	anMethInfo.methSensitivity
anMethInfo.methSpecificity	eMoF	measurementType	anMethInfo.methSpecificity
resVal	eMoF	measurementType	result
resValUnit	eMoF	measurementUnit	result unit

resLOD	eMoF	measurementType	detection limit
resLOQ	eMoF	measurementType	quantification limit
labCountry	eMoF	measurementType	laboratory country location
resID	eMoF	measurementType	test result ID
analysisY	eMoF	measurementDeterminedDate	year
analysisM	eMoF	measurementDeterminedDate	month
analysisD	eMoF	measurementDeterminedDate	day
paramType	eMoF	measurementType	aggregated data
paramCode.anthH	eMoF	measurementType	N antigen
paramCode.antN	eMoF	measurementType	H antigen
anPortSeq	eMoF	measurementType	sequential number
anPortSize	eMoF	measurementType	weight
anPortSizeUnit	eMoF	measurementUnit	
sampStrategy	eMoF	measurementType	sample strategy
sampMatCode.base	eMoF	measurementType	sample material
resType	eMoF	measurementUnit	resType
resUnit	eMoF	measurementUnit	resUnit
anMethType	eMoF	measurementType	result determined date
labID	eMoF	measurementDeterminedBy	laboratory ID
sampUnitIds.subUnitId	eMoF	measurementType	farmID

After examination, we conclude that it is currently possible to map the terms present in the SIGMA database to Darwin Core Standards (Figure 13). The country, surveillance programme and year of collection can be stored in the Event Core (Table 14b). Details on the host individual and the tissue sample can be stored in the Occurrence extension (Table 14c). Laboratory analysis results and disease diagnosis are stored in the eMoF extension, and we make use of nestedness therein to provide details on the laboratory outcome (Table 14d, Figure 13). Information describing either the event, the occurrence, or the tissue sample can also be stored in the eMoF extension (Table 14d). This complex structure can be modelled efficiently in Darwin Core:

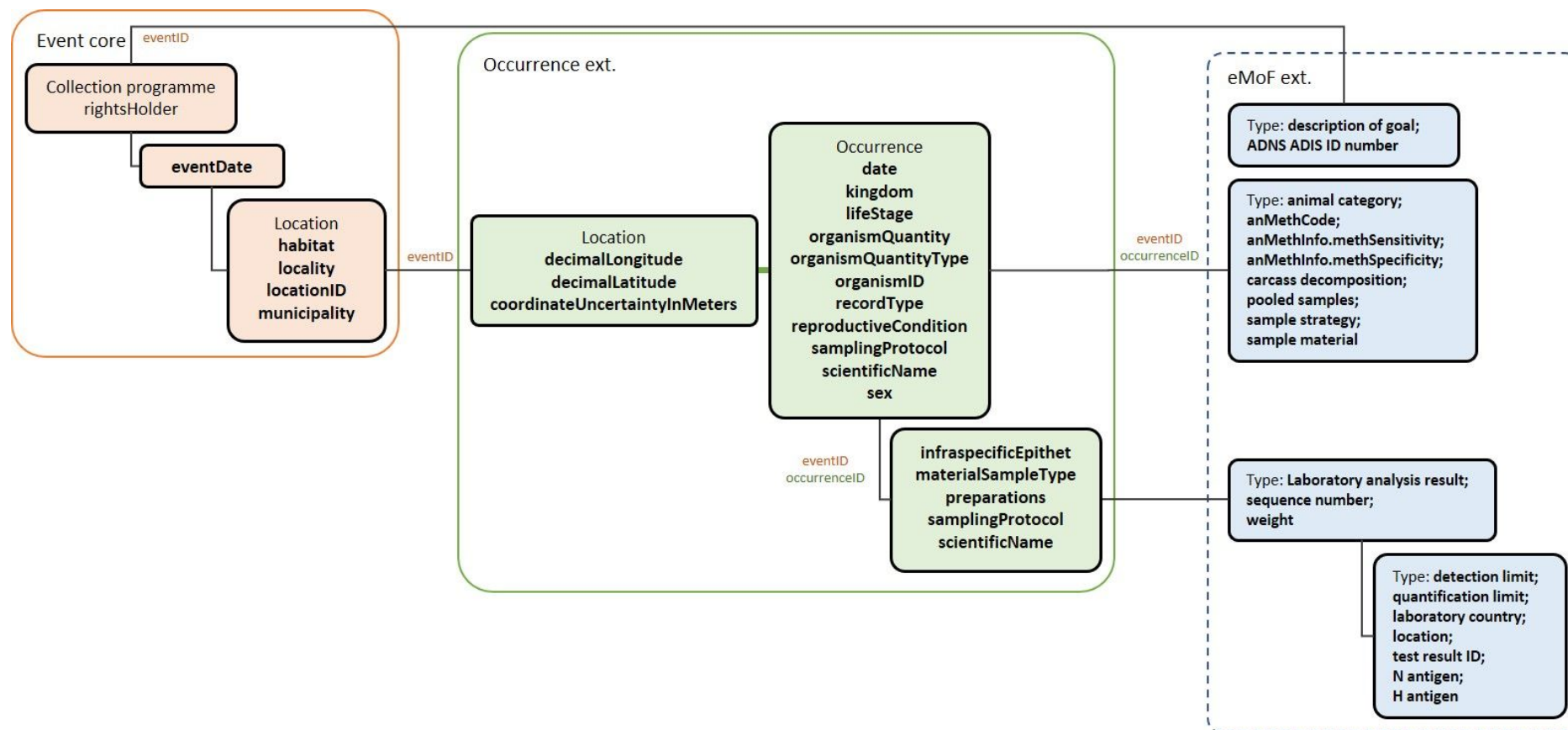


Figure 13: conceptual model for the SIGMA database

Below the corresponding Event (Table 14b). Occurrence (Table 14c) and extended MeasurementOrFact tables are listed (Table 14d).

Table 14b: SIGMA Event Core table

event ID	parent Event ID	collection ID	access Rights	rights Holder	habitat	locality	country	country Code	location ID	location According To	municipality	coordinate Uncertainty In Meters	decimalLatitude	decimalLongitude	event Date	samplingProtocol
event1		collection1	EU programme	Reg EU 2016/429	marsh		France	FR	FRJ12	NUTS3	Gard					protocol1

Table 14c: SIGMA Occurrence extension table

occurrenceID	event ID	parent OccurrenceID	kingdom	organismID	preparations	scientific Name	organism Scope	infraspecific Epithet	occurrenceStatus	date	lifeStage	sex	reproductive Condition	record Type
event1/occ1	event1		animalia			Anser anser	multicellular organism			2022-08-08	adult			found dead
event1/occ1-1	event1	event1/occ1			cloacal swab	Influenza A virus	virus	HPAI	present					

Table 14d: SIGMA eMoF extension table

measurement ID	eventID	occurrence ID	parent MeasurementID	Type	Value	Unit	Method	Determined Date	Determined By
event1/meas1	event1			description of goal	Freedom from disease				
event1/meas2	event1			ADNS ADIS ID number					
event1/meas3	event1			sample strategy					
event1/occ1/meas1	event1	event1/occ1		animal category	farmed/wild				
event1/occ1/meas2	event1	event1/occ1		carcass decomposition					
event1/occ1/meas3	event1	event1/occ1		analysis Method Code					
event1/occ1/meas4	event1	event1/occ1		analysis MethInfo Sensitivity					
event1/occ1/meas5	event1	event1/occ1		analysis MethInfo Specificity					
event1/occ1/meas6	event1	event1/occ1		sample material	non-food				
event1/occ1/meas6	event1	event1/occ1		organismQuantity					
event1/occ1-1/meas1	event1	event1/occ1-1		result	positive negative	individual result Type (i.e binomial)			labId

Wildlife disease data model standards

event1/occ1-1/meas1-1	event1	event1/occ1-1	event1/occ1-1/meas1	result value	resVal	result Unit		date	
event1/occ1-1/meas1-2	event1	event1/occ1-1	event1/occ1-1/meas1	detection limit					
event1/occ1-1/meas1-3	event1	event1/occ1-1	event1/occ1-1/meas1	quantification limit					
event1/occ1-1/meas1-4	event1	event1/occ1-1	event1/occ1-1/meas1	laboratory country location					
event1/occ1-1/meas1-5	event1	event1/occ1-1	event1/occ1-1/meas1	test result ID					
event1/occ1-1/meas1-6	event1	event1/occ1-1	event1/occ1-1/meas1	year					
event1/occ1-1/meas1-7	event1	event1/occ1-1	event1/occ1-1/meas1	month					
event1/occ1-1/meas1-8	event1	event1/occ1-1	event1/occ1-1/meas1	day					
event1/occ1-1/meas1-9	event1	event1/occ1-1	event1/occ1-1/meas1	pooled samples	yes no				
event1/occ1-1/meas1-10	event1	event1/occ1-1	event1/occ1-1/meas1	N antigen					
event1/occ1-1/meas1-11	event1	event1/occ1-1	event1/occ1-1/meas1	H antigen					
event1/occ1-1/meas1-12	event1	event1/occ1-1	event1/occ1-1-1/meas1	sequence number					
event1/occ1-1/meas2-13	event1	event1/occ1-1	event1/occ1-1-1/meas1	weight		unit			

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.4. ADIS database

The EU Animal Disease Information System (ADIS) is a disease management tool, registering all outbreaks of infectious animal diseases listed in the framework of the Animal Health Law². The database is used for ensuring transparent sharing of animal diseases occurrence, which is of particular interest for early warning and animal trade. The scale of the data reported here is at the outbreak scale, not at the individual scale, which creates the difference between the ADIS and the SIGMA databases. Furthermore, ADIS only reports positive cases, whereas SIGMA reports both positive and negative cases.

The database consists of 1 record for each outbreak. Each record contains a unique identification of the outbreak, its location and timing, in addition to details of the outbreak, diagnosis technique and laboratory results, and an official confirmation (Table 15a).

ADIS terms	Hierarchical arrangement	darwinCore terms	controlled vocabulary
User ID	metadata		
WAHIS Login	metadata		
Password	metadata		
Language of the interface	metadata		
Default OIE language	metadata		
Country/territory code	event	countryCode	
Country/territory name	event	country	
Country type	eMoF	measurementType	
LAU code	event	locationID	
LAU name	event	locality	
Animal type	eMoF	measurementType	
Reason for notification	eMoF	measurementType	
The event occurs in	eMoF	measurementType	
Report ID	eMoF	measurementType	
ADIS report ID	eMoF	measurementType	
National report reference	eMoF	measurementType	
Follow up report occurrence number	eMoF	measurementType	
Date of the report	eMoF	measurementType	
Date/time of submission in ADIS of the outbreak	eMoF	measurementType	
Date of start of the event	event	beginDate	
Date of confirmation of the event	eMoF	measurementDeterminedDate	
date of last occurrence	eMoF	measurementType	
End date of the event	event	endDate	
Emerging disease name	Occurrence	scientificName	
Disease name	Occurrence	scientificName	
Disease Type Subtype name (Genotype/serotype/subtype)	Occurrence	infraspecificEpithet	
Subdisease	eMoF	measurementType	
Type of diagnosis (event level)	eMoF	measurementType	
Type of diagnosis (outbreak level)	Occurrence	preparations	
Change type	eMoF	measurementType	
Mortality & morbidity measurement unit	eMoF	measurementType	
Previous animal mortality	eMoF	measurementType	
Previous animal morbidity	eMoF	measurementType	
Current animal mortality	eMoF	measurementType	
Current animal morbidity	eMoF	measurementType	
Zoonotic potential	eMoF	measurementType	
Import external files	eMoF	measurementType	
Outbreak ADIS ID	eMoF	measurementType	
Outbreak Wahis ID	eMoF	measurementType	
outbreak national reference	eMoF	measurementType	

² https://food.ec.europa.eu/animals/animal-diseases/animal-disease-information-system-adis_en

Wildlife disease data model standards

Outbreak ADNS ID	eMoF	measurementType	
Outbreak Type	eMoF	measurementType	
related Outbreak Reference	eMoF	measurementType	
Date of start of the outbreak/Date of suspicion	eMoF	measurementType	
Date of end of the outbreak/Date of preliminary cleaning and disinfection (where applicable)	eMoF	measurementType	
Outbreak Status	eMoF	measurementType	
Date of confirmation of the outbreak	eMoF	measurementType	
administrative division name	eMoF	measurementType	
Region level	event	locationType	
Outbreak cluster	eMoF	measurementType	
Epidemiological Unit Type name	eMoF	measurementType	
Location	event	locality	
Approximate location	Occurrence	coordinateUncertaintyInMeters	
Latitude	Occurrence	decimalLatitude	
Longitude	Occurrence	decimalLongitude	
Category	eMoF	measurementType	
Species name	Occurrence	scientificName	
Livestock production type	eMoF	measurementType	
Type of wildlife	eMoF	measurementType	
Order	Occurrence	order	
Family name	Occurrence	family	
Latin name	Occurrence	scientificName	
Common name	Occurrence	vernacularName	
Farmed or wild population	eMoF	measurementType	
Water	eMoF	measurementType	
Production system name	eMoF	measurementType	
Measuring units	eMoF	measurementType	organismQuantityType
Number of susceptible animals	eMoF	measurementType	organismQuantity
Number of cases	eMoF	measurementType	
Number of dead animals	eMoF	measurementType	
Number of killed animals	eMoF	measurementType	
Number of slaughtered animals	eMoF	measurementType	
Number of vaccinated animals	eMoF	measurementType	
Specific information for the outbreak	eMoF	measurementType	
Control measures applied at event level (dom. vs wild)	eMoF	measurementType	
Control measures to be applied at event level (dom. vs wild)	eMoF	measurementType	
Control measures applied at outbreak level	eMoF	measurementType	
Are there clinical signs?	eMoF	measurementType	
Causal agent	eMoF	measurementType	
Source of infection (at event level)	eMoF	measurementType	
source of infection applied at outbreak level	eMoF	measurementType	
Epidemiological additional Information	eMoF	measurementType	
Test requested (at event level)	eMoF	measurementType	
Name of Laboratory (event level)	eMoF	measurementType	
Type of laboratory name (event level)	eMoF	measurementType	
Species sampled name (outbreak level)	eMoF	measurementType	
Laboratory results date (outbreak level)	eMoF	measurementType	
Laboratory tests results (outbreak level)	eMoF	measurementType	
Mortality & morbidity measurement unit (outbreak level)	eMoF	measurementUnit	
Morbidity (at outbreak level)	eMoF	measurementType	
Mortality (at outbreak level)	eMoF	measurementType	
Final report	eMoF	measurementType	

We conclude that it is currently possible to map the terms present in the ADIS database to Darwin Core Standards (Figure 14). In this example, however, there is heavy reliance on the eMoF extension to store the details on the events and outbreaks, in addition to the reporting and sampling strategies (Table 15d). This complex structure can be modelled efficiently in Darwin Core.

Wildlife disease data model standards

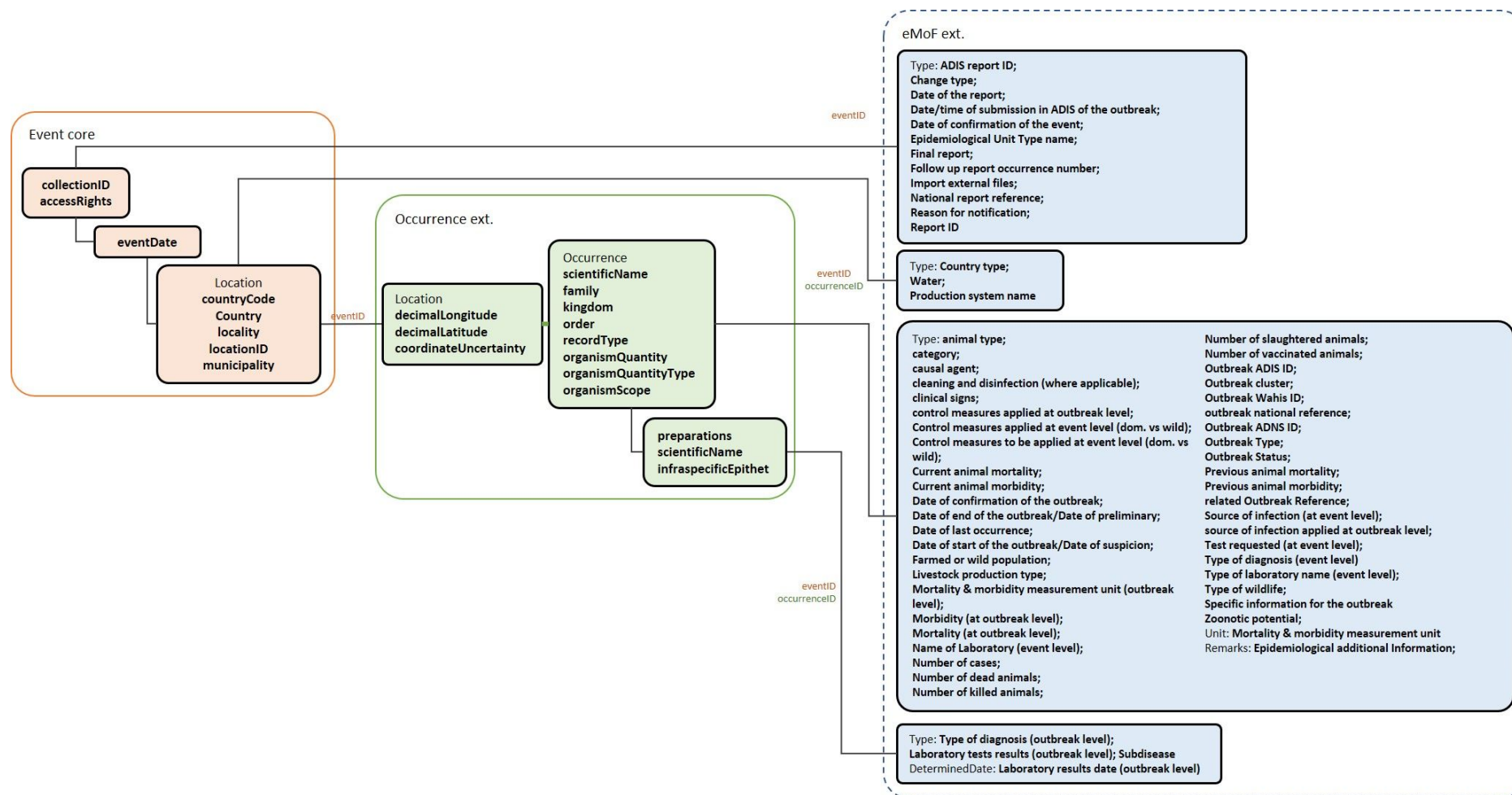


Figure 14: conceptual model for the ADIS database in DwC

Table 15a: Event core

eventID	parent EventID	collectionID	country Code	country	location ID	municipality	eventDate	location Type	locality
event1									
event1-1	event1								
event1-1-1	event1-1		FR	France					

Table 15b: Occurrence table

occurrence ID	event ID	parent OccurrenceID	basisOf Record	decimalLatitude	decimalLongitude	coordinateUncertaintyIn Meters	scientific Name	infraspecific Epithet	order	family	vernacular Name	preparations
event1-1-1/occ1	event1-1-1		human observation				Anser anser		Anseriformes	Anatidae	greylag goose	
event1-1-1/occ1-1	event1-1-1	event1-1-1/occ1	material sample				HPAI	variant				sample type

Table 15c: eMoF extension

measurementID	eventID	occurrenceID	measurementType	measurement Value	measurement Unit	measurementRemarks	measurementDeterminedDate
event1/meas1	event1		Reason for notification				
event1/meas2	event1		The event occurs in				
event1/meas3	event1		Report ID				
event1/meas4	event1		ADIS report ID				
event1/meas5	event1		National report reference				
event1/meas6	event1		Follow up report occurrence number				
event1/meas7	event1		Date of the report				
event1/meas8	event1		Date/time of submission in ADIS of the outbreak				
event1/meas9	event1		Change type				
event1/meas10	event1		Import external files				
event1/meas11	event1		Epidemiological Unit Type name				
event1/meas12	event1		Final report				
event1-1/meas1	event1-1		Date of confirmation of the event				
event1-1-1/meas1	event1-1-1		Country type				
event1-1-1/meas2	event1-1-1		Water				
event1-1-1/meas3	event1-1-1		Production system name				

Wildlife disease data model standards

event1-1-1/occ1/meas1	event1-1-1	event-1-1-1/occ1	Zoonotic potential				
event1-1-1/occ1/meas2	event1-1-1	event-1-1-1/occ1	Outbreak ADIS ID				
event1-1-1/occ1/meas3	event1-1-1	event-1-1-1/occ1	Outbreak Wahis ID				
event1-1-1/occ1/meas4	event1-1-1	event-1-1-1/occ1	outbreak national reference				
event1-1-1/occ1/meas5	event1-1-1	event-1-1-1/occ1	Outbreak ADNS ID				
event1-1-1/occ1/meas6	event1-1-1	event-1-1-1/occ1	Outbreak Type				
event1-1-1/occ1/meas7	event1-1-1	event-1-1-1/occ1	related Outbreak Reference				
event1-1-1/occ1/meas8	event1-1-1	event-1-1-1/occ1	Date of start of the outbreak/Date of suspicion				
event1-1-1/occ1/meas9	event1-1-1	event-1-1-1/occ1	Date of end of the outbreak/Date of preliminary cleaning and disinfection (where applicable)				
event1-1-1/occ1/meas10	event1-1-1	event-1-1-1/occ1	Outbreak Status			Specific information for the outbreak	
event1-1-1/occ1/meas11	event1-1-1	event-1-1-1/occ1	Date of confirmation of the outbreak				
event1-1-1/occ1/meas12	event1-1-1	event-1-1-1/occ1	Outbreak cluster				
event1-1-1/occ1/meas13	event1-1-1	event-1-1-1/occ1	Farmed or wild population				
event1-1-1/occ1/meas14	event1-1-1	event-1-1-1/occ1	Number of cases				
event1-1-1/occ1/meas15	event1-1-1	event-1-1-1/occ1	Number of dead animals				
event1-1-1/occ1/meas16	event1-1-1	event-1-1-1/occ1	Number of killed animals				
event1-1-1/occ1/meas17	event1-1-1	event-1-1-1/occ1	Number of slaughtered animals				
event1-1-1/occ1/meas18	event1-1-1	event-1-1-1/occ1	Number of vaccinated animals				
event1-1-1/occ1/meas19	event1-1-1	event-1-1-1/occ1	Control measures applied at event level (dom. vs wild)				
event1-1-1/occ1/meas20	event1-1-1	event-1-1-1/occ1	Control measures to be applied at event level (dom. vs wild)				
event1-1-1/occ1/meas21	event1-1-1	event-1-1-1/occ1	Control measures applied at outbreak level				
event1-1-1/occ1/meas22	event1-1-1	event-1-1-1/occ1	Causal agent				
event1-1-1/occ1/meas23	event1-1-1	event-1-1-1/occ1	Source of infection (at event level)				
event1-1-1/occ1/meas24	event1-1-1	event-1-1-1/occ1	source of infection applied at outbreak level				
event1-1-1/occ1/meas25	event1-1-1	event-1-1-1/occ1	Epidemiological additional Information				
event1-1-1/occ1/meas26	event1-1-1	event-1-1-1/occ1	Category				
event1-1-1/occ1/meas27	event1-1-1	event-1-1-1/occ1	Livestock production type				
event1-1-1/occ1/meas28	event1-1-1	event-1-1-1/occ1	Type of wildlife				
event1-1-1/occ1/meas29	event1-1-1	event-1-1-1/occ1	Animal type				
event1-1-1/occ1/meas30	event1-1-1	event-1-1-1/occ1	date of last occurrence				
event1-1-1/occ1/meas31	event1-1-1	event-1-1-1/occ1	Morbidity (at outbreak level)		Mortality & morbidity measurement unit		
event1-1-1/occ1-1/meas1	event1-1-1	event-1-1-1/occ1-1	Laboratory tests results (outbreak level)				
event1-1-1/occ1-1/meas2	event1-1-1	event-1-1-1/occ1-1	Type of diagnosis (event level)				Laboratory results date (outbreak level)

Wildlife disease data model standards

event-1-1-1/occ1/meas19	event1-1-1	event-1-1-1/occ1	Mortality (at outbreak level)		Mortality & morbidity measurement unit		
event-1-1-1/occ1/meas19	event1-1-1	event-1-1-1/occ1	Clinical signs				
event-1-1-1/occ1/meas19	event1-1-1	event-1-1-1/occ1	organismQuantity		individual		
event-1-1-1/occ1/meas19	event1-1-1	event-1-1-1/occ1	Previous animal mortality		Mortality & morbidity measurement unit		
event-1-1-1/occ1/meas19	event1-1-1	event-1-1-1/occ1	Previous animal morbidity		Mortality & morbidity measurement unit		
event-1-1-1/occ1/meas19	event1-1-1	event-1-1-1/occ1	Current animal mortality		Mortality & morbidity measurement unit		
event-1-1-1/occ1/meas19	event1-1-1	event-1-1-1/occ1	Current animal morbidity		Mortality & morbidity measurement unit		
event1-1-1/occ1-1/meas1	event1-1-1	event1-1-1/occ1-1	Emerging disease name				
event1-1-1/occ1-1/meas2	event1-1-1	event1-1-1/occ1-1	Type of diagnosis (outbreak level)				
event1-1-1/occ1-1/meas3	event1-1-1	event1-1-1/occ1-1	disease subtype				
event1-1-1/occ1-1/meas4	event1-1-1	event1-1-1/occ1-1	Test requested (at event level)				
event1-1-1/occ1-1/meas5	event1-1-1	event1-1-1/occ1-1	Name of Laboratory (event level)				
event1-1-1/occ1-1/meas6	event1-1-1	event1-1-1/occ1-1	Type of laboratory name (event level)				

3.5. Preparation for Darwin Core developments

There are currently discussions at GBIF and Biodiversity Information Standards, formerly Taxonomic Databases Working Group (<https://www.tdwg.org/>) of substantial developments to their data model. These changes, when they take place, may put more emphasis on collections, and samples preserved in collections. There will likely be no Occurrence extension, all data being transferred to the Event table. Furthermore, there may be more flexibility for dependency *within* tables, such as parentTaxonID and its dependent taxonID.

In anticipation of these developments, we outline below modifications to our proposed Wildlife Disease data model. Figure 15 illustrates two possible solutions to the new developments; firstly, we illustrate that simply nesting within the event core allows us to maintain the integrity of wildlife disease data. That is, that a surveillance programme occurs on a particular year (event), within which multiple locations are surveyed and one or many observations made of a host species at a particular point in time (occurrence) at a particular location (sub-occurrence). The host individual may be the object of a tissue sample (sub sub-occurrence) for laboratory analysis and disease diagnosis (material sample measurement, Figure 15a).

Secondly, an alternative approach that does not rely solely on the Event core but rather aligns with the increased importance of tables such as the materialSamples table that is currently proposed in the new GBIF data model (Figure 15b). Instead of the tissue sample corresponding to a sub-occurrence, the host individual observation will instead be linked to the materialSample table via a materialSampleID. Both the materialSample and the host observation can be associated to measurements or facts associated. This is further illustrated in tables 16a-d.

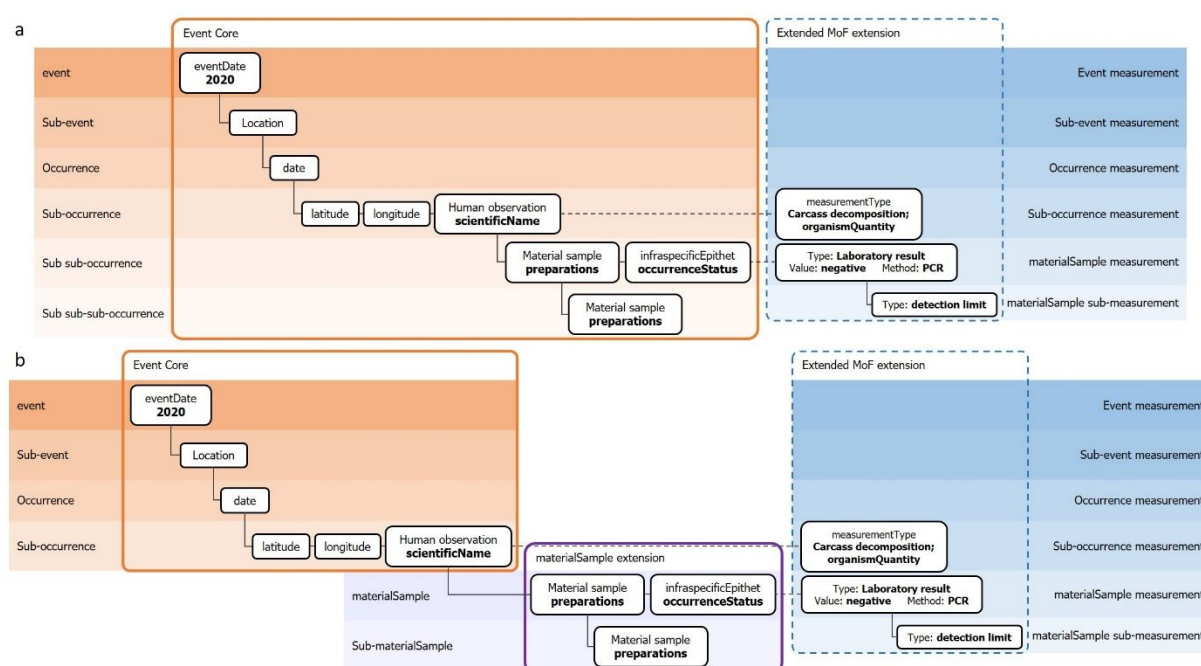


Figure 15: application of nestedness to the Epidemiology Data Model using (a)

One question that remains here is the diversity of the materialSample extension. It is imperative that information pertaining to the species can still be stored in this extension, in order to distinguish between

pathogen variants. This remains uncertain however, and further investigation may be necessary before the new developments in Darwin Core are implemented in Enetwild.

Table 16a: raw data

species	nr observed	location	Date	body temperature deg	blood sample taken	ASFv present
wild boar	1	Auriol	19/03/2020	39.6	yes	yes
wild boar	1	Auriol	22/03/2020		yes	yes
wild boar	1	Cabriès	30/01/2018		yes	no

Table 16b: the Event table for a case of African Swine Fever Virus (ASFv) present in wild boar

eventID	parentEventID	locality	scientific Name	organism Scope	occurrenceStatus	eventDate
event1		Auriol				
event2		Cabriès				
event1-1	event1		Sus scrofa	multicellular organism	present	2020-03-19
event1-2	event1		Sus scrofa	multicellular organism	present	2020-03-22
event2-1	event2		Sus scrofa	multicellular organism	present	2018-01-30

Table 16c: materialSample extension table

material SampleID	eventID	preparations	scientific Name	infraspecific Epithet	organism Scope	occurrence Status
event1-1/mat1	event1	blood sample	ASFv		virus	present
event1-2/mat1	event1	blood sample	ASFv		virus	present
event1-1/mat1	event2	blood sample	ASFv		virus	absent

Table 16d: eMoF extension table

measurementID	eventID	material SampleID	measurement Type	measurement Value	measurement Unit
event1-1/meas1	event1-1		body temperature	39.6	degree Celsius
event1-1/meas2	event1-1		organismQuantity		individual
event1-1/mat1/meas1	event1-1	event1-1/mat1	result	positive	
event1-2/mat1/meas1	event1-2	event1-2/mat1	result	positive	
event2-1/mat1/meas1	event2-1	event2-1/mat1	result	negative	

4. Conclusions and recommendations

We demonstrate that Darwin Core is a data standard capable of sharing epidemiological data, from local scientific surveys, to national and European-scale disease surveillance systems. The ability to transform epidemiological data to Darwin Core has the potential to create a bridge between wildlife population data and epidemiological data at a European scale. Establishing a Darwin Core connection to the SIGMA and ADIS databases (demonstrated in sections 3.3 and 3.4) would open up many possibilities for making disease data publicly accessible. Via the Global Biodiversity Information Facility (GBIF) there is even potential to reach a global scale for data compatibility, given confidentiality issues are respected, by creating a plug-in. Such a translation to the GBIF version of Darwin Core is achievable, as demonstrated in section 3.5, since our propositions for increased nesting versatility is already under consideration for the new GBIF data model. Furthermore, a use case is being developed specifically before the release of the GBIF data model to demonstrate the model capacities for inclusion of epidemiological data.

Further, despite the macro scale of the ADIS database, the ability to standardise these data is nevertheless of interest for associating to wildlife distribution data, since the location and the duration of the outbreak remain present.

We propose a controlled vocabulary for animal disease data that aligns with the terms already listed for the SIGMA and ADIS database. We recommend this adoption to encourage compatibility between local and European-scale data, and broadscale transparency for smoother connections with wildlife distribution data. One of the considerations to bear in mind when using the Darwin Core standard is its flexibility, which is both its strength and its weakness. There are no regulations, only guidelines, on controlled vocabulary within the term fields, meaning that strict discipline is required to maintain unambiguous and interoperable terms in, for example, the *measurementType* field of the eMoF extension. This is particularly crucial in an international context, where language differences can additionally create problems in definition overlaps among vocabulary.

We advocate for hierarchical structure in the Occurrence extension and in the extended MeasurementOrFact extension for modelling cases where Occurrences or Material samples are subsampled or dependent on other Occurrences or Material samples, and when measurements are derived from other measurements. These modifications to Darwin Core would facilitate the storage and sharing of complex biodiversity and epidemiological data, to name just two examples. The proposed data model can be employed as a database structure in its own right, or simply as a front for translation during the import and export phase of existing wildlife disease databases.

This exploration is a clear demonstration of the benefits and uses of the proposed *parentOccurrenceID* and *parentMeasurementID* terms to the Darwin Core standard. If included in the forthcoming developments of Darwin Core standard model, these terms would not only enable the first step toward combining wildlife and disease data, but open possibilities for smoother storage of other complex data sets. Finally, our proposed data model and potential links to facilities such as GBIF demonstrate the power of the Darwin Core standard for unifying efforts on disease surveillance from a local scale to a global scale.

References

1. Consortium, E. *et al.* Applying the Darwin core standard to the monitoring of wildlife species, their management and estimated records. *EFSA Support. Publ.* **17**, 1841E (2020).
2. Cohen Jeremy M., Sauer Erin L., Santiago Olivia, Spencer Samuel, & Rohr Jason R. Divergent impacts of warming weather on wildlife disease risk across climates. *Science* **370**, eabb1702 (2020).
3. Wieczorek, J. *et al.* Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLOS ONE* **7**, e29715 (2012).
4. Desmet, P., Bubnicki, J. & Norton, B. Camtrap DP: A frictionless data exchange format for camera trapping data. *Biodivers. Inf. Sci. Stand.* **5**, e73188 (2021).
5. Groom, Q. *et al.* Improving Darwin Core for research and management of alien species. *Biodivers. Inf. Sci. Stand.* **3**, e38084 (2019).
6. De Pooter, D. *et al.* Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. *Biodivers. Data J.* e10989–e10989 (2017) doi:10.3897/BDJ.5.e10989.
7. Picard-Meyer, E. *et al.* Bat rabies surveillance in France: first report of unusual mortality among serotine bats. *BMC Vet. Res.* **13**, 387 (2017).

Appendix

A.1. List of referenced terms employed for the Epidemiological Enetwild data model

Below are listed the terms employed beyond those proposed by the Enetwild Consortium¹:

A.1.1. Event table

Table	Concept	Variable	Definition	Type	N	Example	DwC	ISO
Event	Identification	datasetID	The identifier of the dataset as found in the metadata	string/doi	1		http://rs.tdwg.org/dwc/terms/datasetID	
		eventID	An identifier for the set of information associated with an Event (something that occurs at a place and time). May be a global unique identifier or an identifier specific to the data set.	string/doi	1	IT001A1000004	http://rs.tdwg.org/dwc/terms/eventID	
		parentEventID	An identifier for the broader Event that groups this and potentially other Event	string/doi	1	IT001A1000004	http://rs.tdwg.org/dwc/terms/parentEventID	
		eventName	The common name of the Event	string	1	Manoria		
	Origin	originalID	the original ID of the dataset	string/doi	1			
		recordedBy	A person, group, or organization responsible for recording the original Event.	string	1	Mario Rossi		
		accessRights	Information about who can access the resource or an indication of its security status.	string	1			
	Location	locality	The specific description of the place. This term may contain	str	1		http://rs.tdwg.org/dwc/terms/locality	

			information modified from the original to correct perceived errors or standardize the description.					
	locationType	Type of delineation of the geographic information	loc_type	1	polygon			
	xyType	Nature of the X,Y values (real or estimated) to be reported	loc_xytype	1	estimated			
	decimalLongitude	Geographic Longitude (in decimal degree, using the spatial reference system in "Reference system")	numeric	1	7.210178	http://rs.tdwg.org/dwc/terms/decimalLongitude		
	decimalLatitude	Geographic Latitude (in decimal degree, using the spatial reference system in "Reference system")	numeric	1	45.520042	http://rs.tdwg.org/dwc/terms/decimalLatitude		
	xyUncertainty	The horizontal distance class (in meters) from the given X and Y describing the smallest circle containing the whole of the Location. If the uncertainty is unknown or cannot be estimated select option NA (not applicable)	loc_uncertainty	1	100m-1km	http://rs.tdwg.org/dwc/terms/coordinateUncertaintyInMeters		
	footprintWKT	A Well-known Text (WKT) representation of the shape that defines the location	string	1	POLYGON((5 5,28 7, 44 14, 47 35,40 40,20 30,5 5))	http://rs.tdwg.org/dwc/terms/footprintWKT		
	locationID	Identifier of the line in the provided shapefile or Code corresponding to the cell of the reference grid to which the data refers or to the NUT identification (see locationAccordingTo)	string	1	ID3526 or 10kmE283N286 or ITI18 - Arezzo	http://rs.tdwg.org/dwc/terms/locationID		

		locationAccordingTo	The name for the shape file provided (assigned by data provider) or to the reference grid or reference NUT	string	1	CR12082.shp or Grid_ETRS89_LAEA_10K_F R or NUTS3-2016	http://rs.tdwg.org/dwc/terms/locationAccordingTo	
		referenceSystem	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the X,Y coordinates or polygon are given	string	1	EPSG: 23030	http://rs.tdwg.org/dwc/terms/geodeticDatum	
		country	Country where data were collected		1	Italy	http://rs.tdwg.org/dwc/terms/country	
		countryCode	The standard code for the country in which the Location occurs.	loc_country	1	IT	http://rs.tdwg.org/dwc/terms/countryCode	
		municipality	The full, unabbreviated name of the next smaller administrative region than county (city, municipality, etc.) in which the Location occurs.				http://rs.tdwg.org/dwc/terms/municipality	
		areaType	Type of the area which data refer.	loc_areatyp	1	hunting ground		
		areaSize	Size of this area	numeric	1	4503		
		areaSizeUnit	Unit of the area	unit_area	1	hectare		
	Time	timeLevel	Level of temporal aggregation of data	time_level	1	day		
		eventDate	The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological context.	ISO 8601-1:2019	1	2009-02-20	http://rs.tdwg.org/dwc/terms/eventDate	
		dateTime	Original representation of the date-time or interval during	string	1		http://rs.tdwg.org/dwc/terms/eventDate	

			which an Event occurred. It can be a precise date-time or a range (e.g., hunting season)				
	Protocol	samplingProtocol	The names of, references to, or descriptions of the methods or protocols used during an Event.	string	1	ad hoc observation point count	http://rs.tdwg.org/dwc/terms/samplingProtocol
	Notes	eventRemarks	notes about the event	string	1		http://rs.tdwg.org/dwc/terms/eventRemarks

A.1.2. Occurrence table

Table	Concept	Variable	Definition	Type	N	Example	DwC	ISO
Occurrence	Identification	datasetID	The identifier of the dataset as found in the metadata	string/doi	1		http://rs.tdwg.org/dwc/terms/datasetID	
		occurrenceID	An identifier for the Occurrence.	string/doi	1	IT001A1000004	http://rs.tdwg.org/dwc/terms/occurrenceID	
		eventID	An identifier for the Event.	string/doi	1	IT001A1000004	http://rs.tdwg.org/dwc/terms/parentEventID	
		parentOccurrenceID	An identifier for the broader Occurrence that groups this and potentially other Occurrences	string	1	Manoria		
	Origin	originalID		string/doi	1			

		recordedBy	A person, group, or organization responsible for recording the original Event.	string	1	Mario Rossi		
	Location	locality	The specific description of the place. This term may contain information modified from the original to correct perceived errors or standardize the description.	str	1		http://rs.tdwg.org/dwc/terms/locality	
		locationType	Type of delineation of the geographic information	loc_type	1	polygon		
		xyType	Nature of the X,Y values (real or estimated) to be reported	loc_xytype	1	estimated		
		decimalLongitude	Geographic Longitude (in decimal degree, using the spatial reference system in "Reference system")	numeric	1	7.210178	http://rs.tdwg.org/dwc/terms/decimalLongitude	
		decimalLatitude	Geographic Latitude (in decimal degree, using the spatial reference system in "Reference system")	numeric	1	45.520042	http://rs.tdwg.org/dwc/terms/decimalLatitude	
		xyUncertainty	The horizontal distance class (in meters) from the given X and Y describing the smallest circle containing the whole of the Location. If the uncertainty is unknown or cannot be estimated select option NA (not applicable)	loc_uncertainty	1	100m-1km	http://rs.tdwg.org/dwc/terms/coordinateUncertaintyInMeters	
		footprintWKT	A Well-known Text (WKT) representation of the shape that defines the location	string	1	POLYGON((5 5,28 7, 44 14, 47 35,40 40,20 30,5 5))	http://rs.tdwg.org/dwc/terms/footprintWKT	
		locationID	Identifier of the line in the provided shapefile or Code corresponding to the cell of the reference grid to which the data	string	1	ID3526 or 10kmE283N286 or ITI18 - Arezzo	http://rs.tdwg.org/dwc/terms/locationID	

			refers or to the NUT identification (see locationAccordingTo)					
		locationAccordingTo	The name for the shape file provided (assigned by data provider) or to the reference grid or reference NUT	string	1	CR12082.shp or Grid_ETRS89_LAEA_10K_F R or NUTS3-2016	http://rs.tdwg.org/dwc/terms/locationAccordingTo	
		referenceSystem	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the X,Y coordinates or polygon are given	string	1	EPSG: 23030	http://rs.tdwg.org/dwc/terms/geodeticDatum	
		country	Country where data were collected		1	Italy	http://rs.tdwg.org/dwc/terms/country	
		countryCode	The standard code for the country in which the Location occurs.	loc_country	1	IT	http://rs.tdwg.org/dwc/terms/countryCode	
		municipality	The full, unabbreviated name of the next smaller administrative region than county (city, municipality, etc.) in which the Location occurs.				http://rs.tdwg.org/dwc/terms/municipality	
		areaType	Type of the area which data refer.	loc_area_type	1	hunting ground		
		areaSize	Size of this area	numeric	1	4503		
		areaSizeUnit	Unit of the area	unit_area	1	hectare		
	Time	timeLevel	Level of temporal aggregation of data	time_level	1	day		
		eventDate	The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological context.	ISO 8601-1:2019	1	2009-02-20	http://rs.tdwg.org/dwc/terms/eventDate	

		dateTime	Original representation of the date-time or interval during which an Event occurred. It can be a precise date-time or a range (e.g., hunting season)	string	1		http://rs.tdwg.org/dwc/terms/eventDate	
	notes	occurrenceRemarks	notes about the Occurrence	string	1		http://rs.tdwg.org/dwc/terms/occurrenceRemarks	
	Biology	basisOfRecord	The specific nature of the data record.	bio_basis	1	Human observation	http://rs.tdwg.org/dwc/terms/basisOfRecord	
		scientificName	The full scientific name as used by the consortium.	bio_species	1	Capreolus capreolus		
		occurrenceStatus	A statement about the presence or absence of a Taxon at a Location.	bio_status	1	present	http://rs.tdwg.org/dwc/terms/occurrenceStatus	
		recordType	Type of sightings recorded in the database	bio_record_type	1	alive	http://rs.tdwg.org/dwc/iri/typeStatus	
		sex	The sex of the individual(s) represented in the Occurrence.	bio_sex	1	female	http://rs.tdwg.org/dwc/terms/sex	
		lifeStage	The age class of the individual(s) at the time the Occurrence was recorded	bio_lifeStage	1	juvenile	http://rs.tdwg.org/dwc/terms/lifeStage	
		individualCount	The number of individuals represented present at the time of the Occurrence	integer	1	5	http://rs.tdwg.org/dwc/terms/individualCount	
		individualID	The identification of the individual	string	1	RD608	http://rs.tdwg.org/dwc/terms/organismID	
		order	The full scientific name of the order in which the taxon is classified.	string	1		http://rs.tdwg.org/dwc/terms/order	
		family	The full scientific name of the family in which the taxon is classified.	string	1		http://rs.tdwg.org/dwc/terms/family	

		infraspecificEpithet	The name of the lowest or terminal infraspecific epithet of the scientificName, excluding any rank designation.	string	1		http://rs.tdwg.org/dwc/terms/infraspecificEpithet	
		preparations	A list (concatenated and separated) of preparations and preservation methods for a specimen.	string	1	DNA extract	http://rs.tdwg.org/dwc/terms/preparations	
		reproductiveCondition	The reproductive condition of the biological individual(s) represented in the Occurrence.	string	1	non-reproductive	http://rs.tdwg.org/dwc/terms/reproductiveCondition	

A.1.3. Extended MeasurementOrFact table

Table	Concept	Variable	Definition	Type	N	Example	DwC	ISO
	identification	datasetID	The identifier of the dataset as found in the metadata.	string/doi	1		http://rs.tdwg.org/dwc/terms/datasetID	
		measurementID	An identifier for the MeasurementOrFact (information pertaining to measurements, facts, characteristics, or assertions).	string/doi	1		http://rs.tdwg.org/dwc/terms/measurementID	
		eventID	An identifier for the broader Event that groups this and potentially other Events.	string/doi	1	IT001A1000004	http://rs.tdwg.org/dwc/terms/parentEventID	
		occurrenceID	An identifier for the broader Occurrence that groups this and potentially other Events.	string/doi	1	IT001A1000004	http://rs.tdwg.org/dwc/terms/occurrenceID	
		parentMeasurementID	An identifier for the broader eMoF that groups this and potentially other Events.	string/doi	1			

Wildlife disease data model standards

	Measurement Or Fact	measurementType	The nature of the measurement, fact, characteristic, or assertion.	nemof_type	1		http://rs.tdwg.org/dwc/terms/measurementType	
		measurementValue	The value of the measurement, fact, characteristic, or assertion.	string/numeric	1		http://rs.tdwg.org/dwc/terms/measurementValue	
		measurementUnit	The units associated with the measurementValue.	string/unit_al	1	mm	http://rs.tdwg.org/dwc/terms/measurementUnit	
		measurementMethod	A description of or reference to (publication, URI) the method or protocol used to determine the measurement, fact, characteristic, or assertion.	string	1	polymerase chain reaction	http://rs.tdwg.org/dwc/terms/measurementMethod	
		measurementDetermined Date	The date on which the MeasurementOrFact was made.	ISO date	1	1963-03-08T14:07-0600	http://rs.tdwg.org/dwc/terms/measurementDeterminedDate	
		measurementDetermined By	A list (concatenated and separated) of names of people, groups, or organizations who determined the value of the MeasurementOrFact.	string	1	Laboratory Anses	http://rs.tdwg.org/dwc/terms/measurementDeterminedBy	
	Notes	measurementRemarks	Comments or notes accompanying the MeasurementOrFact.		1	tip of tail missing	http://rs.tdwg.org/dwc/terms/measurementRemarks	

A.2. Controlled vocabulary

A.2.1. Locations lists

country

VALUE	DEFINITION
AD	Andorra
AL	Albania
AM	Armenia
AT	Austria
AZ	Azerbaijan
BA	Bosnia Herzegovina
BE	Belgium
BG	Bulgaria
BY	Belarus
CH	Switzerland
CY	Cyprus
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GB	United Kingdom of Great Britain and Northern Ireland
GE	Georgia
GR	Greece
HR	Croatia
HU	Hungary
IE	Ireland
IT	Italy
LI	Lichtenstein
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Moldova, Republic of
ME	Montenegro
MK	Macedonia
MT	Malta
NL	Netherlands
NO	Norway
PL	Poland
PT	Portugal
RO	Romania
RS	Serbia
RU	Russia
SE	Sweden
SI	Slovenia
SK	Slovakia
TR	Turkey

UA	Ukraine
XZ	Kosovo

loc_type : type of delineation of the geographic information

VALUE	DEFINITION
coordinates	coordinates X and Y are provided, either in WKT format, using the XY columns or in external file
polygon	a spatial shape is provided, either in WKT format or in external file
line	a spatial line is provided, either in WKT format or in external file
EEA grid	spatial information is a cell code of the EEA grid
UTM grid	spatial information is a cell code of the UTM grid
NUTS	spatial information is a European NUT code

loc_xytyp : The nature of the XY values provided

VALUE	DEFINITION
real	exact coordinates measured
estimated	estimation of the coordinates (for instance, centroid of a municipality)
Centroid Admin	Lat / Lon coordinates of the centroid of the administrative unit
Exact	Exact Lat / Lon coordinates

loc_uncert : description of the type of area described

VALUE	DEFINITION
0-10m	the given X,Y coordinates have an uncertainty lower than 10m
10-50m	the given X,Y coordinates have an uncertainty between 10m and 50m
50-100m	the given X,Y coordinates have an uncertainty between 50m and 100m
100m-1km	the given X,Y coordinates have an uncertainty between 100m and 1000m
>1km	the given X,Y coordinates have an uncertainty higher than 1000m
unknown	the uncertainty is unknown

loc_areatyp : description of the type of area described

VALUE	DEFINITION
administrative unit	administrative area, similar a equal to a NUT level
hunting ground	area where a precise hunt happened
management unit	global area managed
study area	area of the study performed
biogeographical unit	biologically relevant area

A.2.2. Lists for time

time_scale : temporal resolution

VALUE	DEFINITION
year	entries in the dataset refer to annual events, e.g. annual census
month	entries in the dataset refer to monthly events, e.g. hunting bag summarized by month
precise date	entries in the dataset refer to single observations or daily events, e.g. individual sighting
hunting season	entries in the dataset refer to single hunting seasons
other	none of the above options can apply

time_level : level of temporal aggregation of data

VALUE	DEFINITION
day	
month	
year	
interval	begin and end date of the interval has to be provided
hunting season	
pre-birth season	
post-birth season	

A.2.3. List for biological records

scientificName: species list of Enetwild

VALUE	DEFINITION
<i>Cervus elaphus</i>	Red deer
<i>Capreolus capreolus</i>	Roe deer
<i>Dama dama</i>	Fallow deer
<i>Capra ibex</i>	Alpine
<i>Capra hispanica</i>	Iberian ibexes
<i>Ovis aries</i>	Mouflon
<i>Rupicapra rupicapra</i>	Chamois
<i>Rupicapra pyrenaica</i>	Southern chamois
<i>Alces alces</i>	Moose
<i>Rangifer tarandus</i>	Reindeer
<i>Bison bonasus</i>	European Bison
<i>Odocoileus virginianus</i>	White-tailed deer
<i>Cervus elaphus</i> subsp	Elk
<i>Hydropotes inermis</i>	Chinese water deer
<i>Muntiacus reevesi</i>	Muntjac deer
<i>Ovibos moschatus</i>	Musk ox
<i>Ammotragus lervia</i>	Barbary sheep
<i>Canis lupus</i>	Wolf
<i>Lynx lynx</i>	Eurasian lynx

Meles meles	European badger
Procyon lotor	Raccoon
Nyctereutes procyonoides	Raccoon dog
Vulpes vulpes	Red fox
Ursus arctos	Brown bear
Canis aureus	Golden jackal
African swine fever virus (Inf. with)	African swine fever
A.S.F. in domestic pigs	African swine fever in domestic pigs
A.S.F. in wild boar	African swine fever in wild boar
Classical swine fever virus (Inf. with)	Classical swine fever
C.S.F. in domestic pigs	Classical swine fever in domestic pigs
C.S.F. in wild boar	Classical swine fever in wild boar
Highly pathogenic influenza A viruses (Inf. with)(non-poultry including wild birds)(2017-)	HPAI (non-poultry including wild birds)
Mycobacterium tuberculosis complex (Inf. with)(2019-)	Infection with Mycobacterium tuberculosis complex
Low pathogenic avian influenza in wild birds (- 04/2021)	LPAI.WB-ADNS
Low pathogenic avian influenza (poultry)(2006-)	Low pathogenic avian influenza
Rabies virus (Inf. with)	Rabies

occurrenceStatus:

VALUE	DEFINITION
present	
absent	
present-stationnal	present in the whole area
present-inventorial	present somewhere in the area

recordType:

VALUE	DEFINITION
alive	the target of observation was observed alive
dead	the target of observation was found dead
indirect sign	indices of the species was observed
other	must be define in notes
Alive	The animal was ALIVE at the moment of the sample collection
Found dead	The animal was DEAD at the moment of the sample collection
Road kills	The animal was DEAD due to the impact with a means of transport at the moment of the sample collection
Hunted	The animal was shot and killed in the context of a hunting session
Slaughtered	The animal was stunned and killed during a regular slaughtering session in an authorised slaughterhouse, INTENDED for human consumption.
Culled	The animal was stunned and/or killed on site or at a slaughterhouse for urgent ans sanitary matters, including stamping out. NOT INTENDED for human consumption
Alive.Symptomatic	The animal was showing clinical signs at the moment of the sample collection or patognomonic lesions

Alive.Non-symptomatic	The animal had no clinical sign nor patognomonic lesion identifiable at the moment of the sample collection
Dead.Symptomatic	The carcass bore clinical signs at the moment of the sample collection or patognomonic lesions
Dead.Non-symptomatic	The carcass didn't bear clinical sign nor patognomonic lesion identifiable at the moment of the sample collection
Hunted.Symptomatic	Hunted with clinical signs (including injured)
Hunted.Non-symptomatic	Hunted without clinical signs

sex:

VALUE	DEFINITION
male	the sex of the observed individual's is male
female	the sex of the observed individual's is female

lifeStage:

VALUE	DEFINITION
adult	the stage of life of the observed individual's is adult/mature
juvenile	the stage of life of the observed individual's is juvenile/immature
embryo/unhatched egg	An unborn or unhatched offspring in the process of development
juvenile – new born	A newly born juvenile is an individual that can be directly and recently linked to the last breeding episode (i.e. individual at nest/den, like chick/pup/cub/kitten/fawn/...)
juvenile - immature	A juvenile that has not yet reached the sexual maturity, but is not anymore a newly born individual.

basisOfRecord:

VALUE	DEFINITION
human observation	the observation was recorded by a human
machine observation	the observation was recorded by an automatic machine
statistical estimation	the record is the result of a statistical estimation
material sample	the record is the result of a tissue sample extraction

Tech_carcass:

Value	Definition
Fresh stage	The carcass basically does not show signs of decomposition. The time elapsed between the death and the finding / sampling of the carcass is short.
Decayed stage	The typical decomposition processes (bloat, active decay, advanced decay) are ongoing. The time elapsed between the death and the finding / sampling of the carcass is relatively short.
Dry/skeletonised stage	The carcass is essentially reduced to its skeleton and dry skin. The time elapsed between the death and the finding / sampling of the carcass is very long (> 2 years)

Context_description:

Value	Definition
Freedom From Disease	Sampling performed with the specific aim of demonstrating freedom from a given disease / parasite / pathogen / agent
Outbreak containment	Sampling performed following a positive laboratory test result
Detection	Regular sampling activity performed with different aims as monitoring or surveillance.

preparations

Value	Definition
tonsils	
lymph nodes	kidney-shaped organ of the lymphatic system, and the adaptive immune system.
gastro-hepatic lymph nodes	Lymph node situated in the context of the gastro-hepatic ligament
renal lymph nodes	Lymph node situated between the kidney (renal arteria) and the aorta
submandibular lymph nodes	Lymph node situated along the medial side of the mandibular bone
retropharyngeal lymph nodes	
spleen	
liver	
kidney	
lung	
bone tissue	Rigid tissue that constitutes part of the vertebrate skeleton in animals
long bone	The long bones (Ossa longa) are those that are longer than they are wide. They grow primarily by elongation of the diaphysis, with an epiphysis at each end of the growing bone.
short bone	Short bones are designated as those bones that are as wide as they are long. Their primary function is to provide support and stability with little to no movement.
bone marrow	
blood	
blood anticoagulated	Blood kept fluid by addition of EDTA or analogous anti-coagulants
blood clotted	Coagulated blood
Blood serum	
skin	
muscle	
Fat	
Salivary glands	
Not-food animal related matrices	
Heart	
Stomach	

oropharyngeal swab	Collection of secretions from the mouth and the pharynx by means of a swab
cloacal swab	Collection of secretions from the posterior orifice that serves as the only opening of the digestive, urinary and reproductive tracts by means of a swab
trachea	Cartilaginous tube that connects the larynx to the bronchi of the lungs

A.2.4. Statistical lists

Samp_process:

VALUE	DEFINITION
PSD-simple random sampling	samples are randomly selected from the sampling frame with no other consideration
PSD-stratified random sampling	samples are randomly selected from strates, themselves splitting the sampling frame into homogeneous groups
PSD-cluster sampling	samples are randomly selected from clusters. Clusters are mutually homogeneous yet internally heterogeneous groups in the sampling frame
PSD-probability proportional to size sampling	the probability of inclusion of a sample depend on its size value
PSD-systematic sampling	the selection of samples from the sampling frame depends on a fix rule
NPS-self selection	individual answers themselves to a questionnaire based on their own attraction to it.
NPS-judgement sampling	the selection of samples depends on someone estimation/knowledge
NPS-convenience sampling	the selection of samples depends on what is feasible
NPS-quota sampling	the sampled is set up to match pre-defined quota per category
NPS-snowball sampling	
NPS-census	it is estimated that all of the individuals are observed, or all of the samples of a sampling frame are measured
non relevant	the concept of sampling process does not apply to the dataset
unknown	the sampling process is not known
Objective sampling - Simple Random Sampling	A simple random sample is a subset of a statistical population in which each member of the subset has an equal probability of being chosen. A simple random sample is meant to be an unbiased representation of a (target) group
Objective sampling - Clustered Sampling	A "cluster" is a collection of study subjects with one or more characteristics in common. In a cluster sample, the sampling unit is LARGER than the unit of concern - E.g. selection of SOME farms (clusters) in which ALL animals (units of concern) are sampled. Application: estimation of the prevalence AT ANIMAL LEVEL (not the prevalence of infected farms)
Objective sampling - Multi-stage Sampling	Special case of the cluster-sample, in which SOME ANIMALS (units of concern) are selected from each selected farm (cluster)
Selective sampling - Risk Based	Risk Based Sampling makes use of relevant risk factors (having an impact on the presence / absence of the agent that has to be detected) and related relative risk values.

Selective sampling - Biologically driven	The sampling is driven by biological considerations, but not supported by scientific evidences (relative risk missing)
Selective sampling - Sanitary shooting	Killing of animals behaving strangely. The only applicable targetGroup is "Hunted.symptomatic" (add BR)
Selective sampling - Sanitary slaughter	Killing of domestic animals, usually at establishment (farm) level, for sanitary reasons (e.g. emergency slaughter, culling, ...)
Convenient sampling - Testing the game bag at game handling establishment	Samples originating from animals hunted during the hunting season, on a voluntary basis by the authorised hunters
Convenient sampling - Active hunting of target species in response to an outbreak	Samples originating from hunting activities promoted by the national or local authorities for disease control. Note, this is NOT a sanitary shooting as the animals may be healthy.
Convenient sampling - Accidental sampling	The samples are obtained without a pre-defined sampling calculation or brought to the laboratory.

Samp_availability:

VALUE	DEFINITION
open online	the sampling frame can be found online and is accessible and usable
open on request	the sampling frame can be provided on demand
restricted	the sampling frame is available but its accessibility is restricted (for instance due to EU general data protection regulation)
non available	the sampling frame cannot be provided
unknown	the sampling frame is unknown

Samp_prot:

VALUE	DEFINITION
counting	a number of individuals/contacts detected is recorded on field
inventory	the list of present species of a community on a location is established
presence-only	the presence of a species is recorded but there are no information but absence of data does not inform on the absence of the species
presence-absence	both presence and absence of the species are recorded
individual identification	individuals recorded are uniquely identified (for instance in a CMR protocol, a telemetry protocol, or while ringing birds)
expert knowledge	the record corresponds to the knowledge of someone, based on their own experience and direct and indirect observations.
questionnaire	the record come from the answer to a questionnaire

Var_out:

VALUE	DEFINITION
density	number of individuals (or other unit) per surface unit
population size	total number of individuals
relative abundance	index of abundance, used for trends
hunting bag size	size of hunting bag
presence probability	the analysis provides a probability of presence of the species, which is not a proof of presence, nor a proof of absence
habitat suitability	the analysis provides a suitability value based on habitat characteristics and habitat selection behaviour of the species
detection probability	the analysis provides (usually in combination with other values) a detection probability
occurrence	the dataset correspond to simple occurrence data
count	the dataset correspond to records of count of individuals/contacts or other units

Ana_fam:

VALUE	DEFINITION
Capture Mark Recapture	
Distance sampling	
Species distribution model	(MaxEnt, Support vector machine, Random forest, Bioclim...)
Habitat suitability	
Regression	
Survey sampling	ex : Horvitz Thompson estimator, hunting bag survey
Camera trapping	
Expert classification	
Multivariate similarity surface	

Inference:

VALUE	DEFINITION
Design-based inference	
Design-based inference – Model assisted	
Model-based inference – Frequentist	
Model-based inference – Bayesian	
Hybrid model/design-based inference	
non relevant	
unknown	

Ana_uncert:

VALUE	DEFINITION
none	

statistical – measure	
statistical – distribution	
statistical – quantile	the distribution is displayed through quantiles
minimal-maximal values	
ad-hoc	
ad-hoc –advanced statistics	
non relevant	

Ana_val:

VALUE	DEFINITION
none	
Cross validation	
Mean square error	
Pearson's correlation	
Area Under the Curve	
Comparison with sightings data	
ad-hoc	
ad-hoc –advanced statistics	
non relevant	

A.2.5. Results lists

Res_qual:

Value	Definition
Positive/Present	The pathogen / the genetic material / the antigens / the antibodies are found in the sample.
Negative/Absent	The pathogen / the genetic material / the antigens / the antibodies are NOT found in the sample.
Inconclusive	

A.2.6. Analysis lists

Ana_meth

Value	Definition
Polymerase chain reaction (PCR)	
Quantitative polymerase chain reaction (QPCR)	
Enzyme-linked immunosorbent assay (ELISA)	
Immunoblotting (IB)	
Immunoperoxidase test (IPT)	
Genotyping	
Virus isolation	
Direct fluorescent antibody test (FAT)	

Ana_metTyp

Value	Definition
Screening	The analytical method does not have the necessary characteristics of a confirmation test OR it was used in a sampling plan with the goal of screening the target population
Confirmation	The analytical method has the necessary characteristics of a confirmation test OR it was used in a sampling plan with the goal of confirming a previous preliminary result

Ana_nantigen

Value	Definition
N-antigen unknown	
N1	
N2	
N3	
N4	
N5	
N6	
N7	
N8	
N9	

Ana_hantigen

Value	Definition
H antigen	
Hrough	
HNT	
H-antigen unknown	
H56	
H55	
H54	
H53	
H52	
H51	
H49	
H48	
H47	
H46	
H45	
H44	
H43	
H42	
H41	
H40	

H39	
H38	
H37	
H36	
H35	
H34	
H33	
H32	
H31	
H30	
H29	
H28	
H27	
H26	
H25	
H24	
H23	
H21	
H20	
H19	
H18	
H17	
H16	
H15	
H14	
H12	
H11	
H10	
H9	
H8	
H7	
H6	
H5	
H4	
H3	
H2	
H1	
H-	
Non-H5	
Non-H7	
Non-H5, Non-H7	

A.2.7. Measurements list

	Group	Value	Accepted type of values	Accepted unit	Definition
ev ent	context	habitat	string		
		weather	string		

		surface	numeric	unit_surface	
		presence of quotas	yes/no		
		description of goal	context_description		
		generic effort	effort distance	numeric	unit_length
			effort surface	numeric	unit_surface
			effort visit	numeric	string
			effort time	numeric	unit_time
			total effort	numeric	string
		sampling	sample weight	numeric	
			sample size	numeric	string
			planned sample size	numeric	string
			sample material	farmed/wild	
	Oc cur rence	technical information	distance	numeric	unit_length
			perpendicular distance	numeric	unit_length
			angle	numeric	unit_angle
			weight	numeric	unit_mass
			cause of death	tech_dea	
			carcass status	tech_carcass	
		estimation type	density	numeric	
			relative abundance	numeric	
			population size	numeric	string
			hunting bag size	numeric	string
eM oF- 1	linked to result		detection probability	numeric	string
			reproductive rate	numeric	string
			survival	numeric	string
			sex-ratio	numeric	string
	result information	qualitative laboratory result	res_qual		
	linked to estimation type	analysis method	ana_meth		
		analysis method type	ana_metTyp		
		N-antigen	ana_nantigen		
		H-antigen	ana_hantigen		
		interval	numeric numeric	see estimation	
		distribution	ana_distr		
		standard deviation	numeric	see estimation	
		variance	numeric	see estimation	
eM oF- 2	linked to interval	median	numeric	see estimation	
		quantile	ana_quant		
	linked to interval	confidence level	numeric	see estimation	
	linked to distribution	n	numeric		
		p	numeric		
		λ	numeric		
		M	numeric		
		N	numeric		
		c	numeric		
		p	numeric		
		μ	numeric		
		σ	numeric		
		α	numeric		
		β	numeric		
		a	numeric		
		b	numeric		

	linked to quantile	θ	numeric		
		k	numeric		
		x_0.025	numeric	see estimation	
		x_0.12	numeric	see estimation	
		x_0.215	numeric	see estimation	
		x_0.31	numeric	see estimation	
		x_0.405	numeric	see estimation	
		x_0.5	numeric	see estimation	
		x_0.595	numeric	see estimation	
		x_0.69	numeric	see estimation	
		x_0.785	numeric	see estimation	
		x_0.88	numeric	see estimation	
		x_0.975	numeric	see estimation	

A.2.8. Units list

VALUE	DEFINITION
square meter	
square kilometer	
hectare	
meter	
kilometer	
second	
minute	
hour	
day	
month	
year	
degree	
radian	
gram	
kilogram	
individual	
individual per kilometer	
individual per square kilometer	
group	
contact	
couple	
degree celsius	
copies per gram	
millilitre	